



Cross-Temporal Meta-Analysis: A Conceptual and Empirical Critique

Cort W. Rudolph¹ · David P. Costanza² · Charlotte Wright² · Hannes Zacher³

Published online: 27 November 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The proper estimation of age, period, and cohort (APC) effects is a pervasive concern for the study of a variety of psychological and social phenomena, inside and outside of organizations. One analytic technique that has been used to estimate APC effects is cross-temporal meta-analysis (CTMA). Although CTMA has some appealing qualities (e.g., ease of interpretability), it has also been criticized on theoretical and methodological grounds. Furthermore, CTMA makes strong assumptions about the nature and operation of cohort effects relative to age and period effects that have not been empirically tested. Accordingly, the goal of this paper is to explore CTMA, its history, and these assumptions. Using a Monte Carlo study, we demonstrate that, in many cases, cohort effects are misestimated (i.e., systematically over- or underestimated) by CTMA. This work provides further evidence that APC effects pose intractable problems for research questions where APC effects are of interest.

Keywords Age · Cohorts · Cross-temporal meta-analysis · Generations · Lifespan

Across literatures, a great deal of effort has been invested to properly estimate the independent influence of age, period, and cohort (APC) effects (e.g., Baltes, 1968; Costanza & Finkelstein, 2015; Costanza, Darrow, Yost, & Severt, 2017; Hofer & Sliwinski, 2006; Koslowski, 1986; Schaie, 1986; Yang, 2008; Yang & Land, 2013). For example, research has looked at the effect of age as a marker of individual development in the work context (e.g., Rudolph & Baltes, 2016), of major historical events on various populations (e.g., Elder, 1974; Elder & Liker, 1982), and on differences among cohorts (e.g., Gerstorf, Ram, Hoppmann, Willis, & Schaie, 2011; Schaie, 2013). However, in most of these studies, age, period, and cohort are not independent of each other and, hence, separating out their relative effects is challenging, if not impossible.

It is understood that, statistically, the separation of age, period, and cohort effects generally represents an intractable problem (Glenn, 1976; Glenn, 2005; Bell & Jones, 2013, 2014). This problem becomes intractable when these three effects are each defined by non-independent temporal variables. Although this has long been acknowledged by statisticians, there is no shortage of theories that beg for the separation of such influences (e.g., meta-theoretical perspectives, such as lifespan psychology, see Rudolph, 2016; theories of personality development, see Roberts & Wood, 2006). Consequently, there have been numerous attempts to address this problem through both statistical and methodological means (Palmore, 1978; Yang & Land, 2006).

One methodology for studying cohort effects, cross-temporal meta-analysis (CTMA; Twenge, 2000) has gained popularity in the personality and individual differences literature. Although the IO/OB/HR research has yet to extensively adopt this methodology (see Bubany & Hansen, 2011; Huang, 2018; Wegman, Hoffman, Carter, Twenge, & Guenole, 2018), evidence from CTMAs has informed research and practice in organizations (see Rudolph and Zacher, 2017, for a review). Yet, the assumptions underlying CTMA and the boundaries for its capacity to draw valid inferences regarding cohort effects have never been tested empirically. Thus, it is not clear how well this technique actually fares with respect to its ability to accurately estimate the parameters that it purports to represent. Establishing this is an important step to take before CTMA is adopted more widely as an accepted methodology in IO/OB/HR.

A pre-print version of this work can be found here: <https://psyarxiv.com/exskp/>. Code to replicate the simulations presented here can be found at: <https://osf.io/mak6y/>. A Shiny web-app is also available here: https://cortrudolph.shinyapps.io/CTMA_Simulation_.

✉ Cort W. Rudolph
cort.rudolph@health.slu.edu

¹ Department of Psychology, Saint Louis University, St. Louis, MO, USA

² The George Washington University, Washington, DC, USA

³ Institute of Psychology, Leipzig University, Leipzig, Germany

We have two objectives in terms of trying to address this gap. First, there has never been a detailed exposition of the methodology required to conduct CTMA. Thus, we summarized and synthesized knowledge regarding this approach by reviewing nearly 30 years of research that has applied CTMA to address questions related to age, period, and cohort effects. Second, given the popularity of CTMA, we sought to unpack its theoretical and methodological assumptions via an empirical test of its core assumptions.

This study contributes to the literature in several ways. With respect to research, we offer a critical perspective on CTMA from both theoretical and empirical lenses. Theoretically, we call into question several of the strong assumptions (i.e., those requiring a high degree of judgment or logical inference) that the CTMA methodology makes about the operation of age, period, and cohort effects. Empirically, we test the core statistical assumption of CTMA via Monte Carlo simulations. Practically, our results call into question conclusions drawn from CTMAs. To this end, such conclusions have been interpreted as evidence for the existence of so-called generational differences—a ubiquitous concept in popular business and management literatures (Rauvola, Rudolph, & Zacher, 2018). More broadly, such conclusions have likewise been implicated in the (re)development of social policies (e.g., education; see Twenge, 2009). Given our findings, questions about the validity of conclusions from CTMAs should be taken seriously, as should their practical relevance. Thus, our study also contributes to an emerging literature that takes a critical perspective on the notion of generational differences (e.g., Rudolph & Zacher, 2017).

Disentangling Age, Period, and Cohort Effects

Before continuing, it is first important to clarify what is meant by age, period, and cohort effects. In general, *age effects* refer to biological or social differences attributable to physical or psychological maturation, life stage, or development that occur regardless of when someone was born. For example, cognitive development occurs across the lifespan and impacts cognitive processes, such as decision-making (Tymula, Belmaker, Ruderman, Glimcher, & Levy, 2013), impulse control (Green, Fry, & Myerson, 1994), and the self-regulation of emotions (Labouvie-Vief, Hakim-Larson, DeVoe, & Schoeberlein, 1989).

Period effects refer to the influence of contemporaneous time and, thus, reflect variation among individuals based on the impact of historical events that affect people across ages—although not necessarily in exactly the same way. For example, the Great Depression had effects on almost everyone in the United States, impacting both younger and older adults. Finally, *cohort effects* refer to differences between groups of individuals that are attributable to their membership in that group. For example, cohorts

can be defined by birth year (e.g., people born in 1964) or by events shared by a specific group of people, such as Puerto Rican survivors of hurricanes Maria and Irma or the women who worked in factories during World War II.

It is also important to differentiate birth cohorts, commonly used in the lifespan literature, from generational cohorts used in the generations literature. Birth cohorts are determined by one's birth year. Generational cohorts, as Rudolph and Zacher (2017) have noted, “represent groupings of birth cohorts that have some meaning attached to them, whereas any given birth cohort by itself can be thought of as a value-free (i.e., decontextualized) generation (e.g., the cohort of people born in 2015)” (p. 3). Following a similar logic, Costanza et al. (2012) defined generational cohorts as “a group of individuals, who are roughly the same age, and who experience and are influenced by the same set of significant historical events during key developmental periods in their lives, typically late childhood, adolescence, and early adulthood” (p. 377).

The issue of appropriately partitioning APC effects is of concern for any research that wishes to study the influence of age (e.g., developmental psychology), contemporaneous contextual influences (e.g., economics), or birth histories (e.g., life course sociology) in isolation. However, researching APC effects is problematic when any of these three sources are studied jointly (i.e., when any two sources are, by necessity, used to define the third source). For example, proper estimation of APC effects is particularly of concern for the study of generations because they are wholly determined by the other two factors, age and period.

As an example, for a study conducted in 2018 (i.e., period) with a sample of 18-year-olds (i.e., age), the birth cohort is fixed (i.e., 2000). This dependency makes fully identifying the variance attributable to any one of the effects impossible (e.g., Bell & Jones, 2013, 2014; Glenn, 1976). To deal with the inseparability of APC effects, it is very common for studies of generational effects to artificially “lump together” ranges of birth years (e.g., those born between 1980 and 1985) to form generational groups. Some recent work has argued that researchers need to think about generations as “fuzzy social constructs” (Campbell, Twenge, & Campbell, 2017, p. 130) instead of as distinct cohorts. Others have argued that we need to reframe generations, and instead think about them as perceptual (e.g., Perry, Golom, Catenacci, Ingraham, Covais, & Molina, 2017), contextual (Urick, Hollensbe, Masterson, & Lyons, 2017), identity-based (Lyons & Schweitzer, 2017), or socially constructed (Rudolph & Zacher, 2015) phenomena.

The Empirical Study of Generations

Despite the computational difficulty in separating APC effects, researchers have nonetheless used a variety of analytical techniques in an attempt to study generational differences. Recently,

Costanza et al. (2017) demonstrated that the three most common analytic techniques used in generations research (i.e., cross-sectional ANOVAs, cross-temporal meta-analysis, and cross-classified hierarchical linear modeling) produce different results when applied to the same data. They found that a specific generation might appear to be higher, lower, or not at all different on several outcome variables based on the analytical technique employed (see Costanza et al., 2017, for a full description of the analytic techniques and results). That is, the identification of generational effects is partially dependent on the analytic technique used.

Cross-Temporal Meta-Analysis

While some have reviewed the relative merits and limitations of these analytic approaches (Costanza et al., 2017; Rudolph & Zacher, 2017; Trzesniewski & Donnellan, 2010), others have suggested that cross-temporal meta-analysis (CTMA) is an appropriate technique to estimate cohort effects. The basic contention is that CTMA allows for the control of age effects and assumes that cohort, not period, effects are the more likely and plausible cause of any observed differences (e.g., Twenge et al., 2008). Although CTMAs have acknowledged this confounding as a distinct liability, they have simultaneously offered that explanations for cohort effects are preferred over equally-likely period effects (e.g., “This study also cannot determine whether the change in narcissism is a purely generational effect or a time-period effect ... it seems likely that much of the shift is a generational rather than a time-period effect”; Twenge et al., 2008, p. 894).

CTMA does have interesting and potentially appealing characteristics, including its ability to make predictions about the magnitude of effect sizes, its basis in traditional meta-analysis, the possibility to estimate true effects sizes in the population, and that it can be easily explained to academic and general audiences. Twenge and colleagues have used CTMA in a number of studies of generational differences (e.g., Twenge & Campbell, 2001; Twenge et al., 2008). They have argued that the technique is appropriate for studying cohort effects, for example stating that the “method allows for the simultaneous examination of age and cohort effects and thus permits us to examine age difference ... while controlling for birth cohort and to examine birth cohort differences within age groups” (Twenge & Campbell, 2001, p. 322). According to Twenge (1997b, p. 38), CTMA answers a call from Gergen (1973), who argued that social psychology ought to provide a complete picture of “historical circumstances and changes across time.”

CTMA was first used to study cohort effects in 1997 (Twenge, 1997a; Twenge, 1997b); however, it was not called CTMA then. Instead, Twenge referred to it as, “a time-lag study of cohort differences” (Twenge, 1997a, p. 308). The technique appears to be first referred to as CTMA by name

in 2000 (Twenge, 2000), a paper in which three of Twenge’s previous articles are cited for the method (Twenge, 1997a; Twenge, 1997b; Twenge, 2001¹). A more-complete description of CTMA appeared in Twenge and Campbell (2001), and the technique is also discussed relative to other analytic approaches that have been used in the generations literature in the paper by Costanza et al. (2017).

History and Background of CTMA

Before exploring the merits of CTMA, it is important to start with a brief history of its foundation, grounding, and use. The statistical roots of CTMA arose from traditional meta-analysis techniques (e.g., Rosenthal & DiMatteo, 2001). The procedures for CTMA generally do not differ from those of traditional meta-analysis (Twenge & Campbell, 2001). Criteria for inclusion are established, primary studies looking at the variable(s) and group(s) of interest are identified, and the relevant statistics (e.g., means, standard deviations) as well as sample characteristics (e.g., sample sizes) are extracted. With meta-analysis, population effect sizes can be estimated by compiling the results of multiple studies based on the idea that the limitations of individual studies can be accounted for to draw more accurate conclusions about this parameter.

However, CTMA detours from traditional meta-analysis in two distinct ways. First, the units of analysis in CTMA are generally cohort means (e.g., the mean level of narcissism among college students year-by-year; see Twenge et al., 2008). That is, the assessment of group differences is based on comparisons of the same or (more typically) similarly aged subjects over time (e.g., ~20-year-olds in 1995, ~20-year-olds in 2005, and ~20-year-olds in 2015) rather than on multiple groups at one point in time (20, 30, and 40-year-olds in 2015). The sample means for the group of interest in each study are obtained and then weighted by sample size (Twenge & Campbell, 2001). Generally speaking, studies with a larger sample size are thought of as being more accurate estimates of the population mean (i.e., they exhibit lower sampling variance) and hence should be given more weight in the CTMA.

Early on when discussing this method, Twenge (1997a, 1997b) cited Hedges and Becker (1986) for a set of “special weighting procedures” (Twenge, 1997b, p. 311; i.e., inverse variance weighting) to calculate the effect size estimates. More specifically, Hedges and Becker (1986) proposed a method for combining estimates of the effect sizes from multiple studies that takes into account the variance associated with each estimate. The procedure involves weighting each estimate by the inverse of its variance (i.e., $1/\sigma^2$). This allows

¹ In Twenge’s (2000) paper, Twenge and Campbell (2000) is cited as an unpublished manuscript. Twenge and Campbell (2000) was later published in 2001 and hence appears as such in the “References” section. This explains the incongruity of citing a later paper (2001) that appeared in an earlier one (2000).

studies that are larger in sample size to carry more weight in the computation of the overall, meta-analytic effect. Because variance is a function of sample size, in practice, inverse variance weighting and sample size weighting often yield functionally equivalent results (Hunter & Schmidt, 2004).

The second detour from traditional meta-analysis is that the overall effect size is calculated using the change in predicted group means over time divided by the pooled standard deviation from each study. Twenge (1997b) cited Wolf (1986) as an example using a similar approach. Specifically, the effect size for the change in the variable of interest over time is computed by dividing the absolute value of the difference between the predicted group means (e.g., the model predicted mean of variable X for 20-year-olds in 1995 *minus* the model predicted mean of variable X for 20-year-olds in 2015, divided by the average within-study standard deviation). This results in a “standardized, scale-invariant estimate” of the effect size (Wolf, 1986, p. 25).

The outcome of interest for CTMA is the relationship between the mean score on the measure used and the year those scores were collected. Mackenzie, Erickson, Deane, and Wright (2014) describe CTMA as a “bivariate correlation between mean scores and years” (p. 12) that “permits a numerical index of the degree to which scores on a measure of interest have changed over time” (p. 13). Twenge et al. (2008) described it similarly, writing that CTMA allows for the analysis of “how ... scores change(d) over time, primarily by examining correlations between mean scores and year of data collection” (p. 881). A positive (or negative) correlation suggests that as time passes, the mean score on the variable of interest and for the group of interest has increased (or decreased).

Because mean scores are weighted (i.e., either by sample size or the inverse of sample variances), this correlation is often reported as a standardized regression coefficient (β_{xy}) from a weighted regression model. For example, Twenge et al. (2008) reported that the relationship between Narcissistic Personality Index (NPI) scores and time among American college students was $\beta_{xy} = 0.53$, concluding that “more recent generations report more narcissistic traits” (p. 883). Of note, regardless of the weighted strategy employed, in the case where there is only one predictor entered into such a model, $\beta_{xy} = r_{xy}$.

In summary, CTMA was originally modified from traditional meta-analysis as a way to compare subjects of the same or similar ages at different points in time on a given trait or measure. Since its inception, CTMA has primarily been used as an attempt to demonstrate cohort effects controlling for age by holding it more-or-less constant. We say “more-or-less” and characterize “same or similar ages,” because CTMA typically relies on age ranges (e.g., samples of college students, Twenge, 1997a, 1997b; samples of children, Twenge, 2000; samples of high school and college students, Twenge & Campbell, 2001), rather than single age cohorts (e.g., 18-year olds). In such a case, age is not being controlled for,

insomuch as there is real age-related variability that is being ignored (e.g., studying samples of college students, but assuming that any age-related variability in college student samples is irrelevant to the manifestation or estimation of cohort or period effects). This is one strong assumption of CTMA, to which we will return later.

As an example of the rapid growth in the acceptance of CTMA, one study, *The Age of Anxiety? Birth Cohort Change in Anxiety and Neuroticism, 1952–1993* (Twenge, 2000), has been cited over 960 times according to Google Scholar and over 330 times according to the Social Science Citation Index. Beyond the work of Twenge and colleagues, studies have used CTMA to examine change over time, including differences between cohorts in intergroup relations (Tan, Huedo-Medina, Lennon, White, & Johnson, 2010), anxiety among Chinese military members (Yang, Cao, Lu, Zhu, & Miao, 2014), general mental ability (Pietschnig, Voracek, & Formann, 2010), body dissatisfaction (Karazsia, Tylka, & Murnen, 2017), justifying inequality (Malahy, Rubinlicht, & Kaiser, 2009), attitudes towards seeking mental health services (Mackenzie et al., 2014), perceived job characteristics (Wegman et al., 2018), and loneliness (Clark, Loxton, & Tobin, 2015).

Criticisms of CTMA

Despite its use and seemingly widespread acceptance, CTMA has been criticized on both methodological (Donnellan, Trzesniewski, & Robins, 2009; Trzesniewski, Donnellan, & Robins, 2008) and theoretical grounds (Rudolph & Zacher, 2017). Regarding methodological critiques, Trzesniewski et al. (2008) and Donnellan et al. (2009) offer that the CTMA methodology relies upon ecological relationships—the results of correlating two variables that each represent individual-level phenomena that exist at an aggregated level of analysis (see Ostroff & Harrison, 1999; Robinson, 1950; Rosenthal, Rosnow, & Rubin, 2000). For CTMA, this is the correlation between the year of the study and the average level of the outcome being investigated.

Beyond the possibility of incorrect interpretations (i.e., the ecological fallacy—inferring individual-level relationships from aggregated data), ecological correlations are often of a greater magnitude than corresponding (i.e., disaggregated) individual-level relationships (e.g., Ostroff, 1993), and they ignore the relative size of individual-level effects within their relative groupings (e.g., so-called “frog pond” relationships; Klein, Danserau, & Hall, 1994). Moreover, the possibility of incorrect assumptions about the directionality and magnitude of an effect (e.g., the so-coined Simpson’s paradox; Blyth, 1972) is not well accounted for by such group-level relationships.

Twenge et al. (2008) have argued that the procedure for deriving the CTMA effect size avoids the ecological fallacy because it involves a transformation wherein the difference between the predicted value for the first year and the predicted

value for the last year is divided by the average standard deviation of the individual samples. At the same time, Twenge has also acknowledged that "... this technique probably still results in somewhat higher effect sizes" (Twenge, Zhang, & Im, 2004, p. 314). Moreover, it has never really been made clear why this transformation makes these estimates interpretable as if they were individual-level relationships. To our knowledge, there is no mathematical reason why such a transformation addresses issues associated with the ecological nature of such relationships.

Assumptions of CTMA

CTMA makes two strong assumptions about the nature and operation of cohort effects relative to age and period effects that are questionable. First, the assumption that mean differences across time are attributable to cohort rather than period effects (e.g., Twenge et al., 2008; Gentile, Wood, Twenge, Hoffman, & Campbell, 2015). Second, a statistical assumption that cohort effects manifest as group mean-level changes (e.g., Twenge, 2008). The former assumption largely concerns the logic of CTMA (see Rudolph & Zacher, 2017, for a critical perspective), which is difficult to refute empirically (see Bianchi, 2014). The latter assumption is largely grounded in the inferences drawn from CTMA, and we can test this assumption empirically.

Logical Assumptions of CTMA

CTMA assumes that after holding chronological age more-or-less constant, birth cohort effects are stronger than period effects and, therefore, any year-by-outcome effects are more plausibly attributable to cohort membership than contemporaneous period influences. Twenge makes this argument in a number of papers, including Twenge et al. (2008) when writing about narcissism: "Given the relative stability of social dominance after young adulthood [...], as well as cross-sectional research showing lower narcissism scores in older adults [...], it seems likely that much of the shift is a generational rather than a time-period effect" (p. 894). While this assertion is hard to disprove, there is research that offers two main arguments against it.

First, this argument rests on the largely untenable idea that developmental experiences early during one's lifespan are more important for the formation and ratification of stable individual difference traits (e.g., personality) than later life experiences. As suggested by Rudolph and Zacher (2017), "... research concerning generations has assumed that there is diminishing intraindividual variability in traits over time (i.e., increased rigidity and decreased plasticity) associated with the process of aging. Such "crystallization" results in decreasing age-graded differential susceptibility to contextual influences" (p. 3). However, there is evidence that many trait-

like individual differences develop and change across the lifespan. For example, work by Roberts, Walton, and Viechtbauer (2006) has demonstrated that there are distinct patterns of mean-level changes in personality traits across the life course. Similar results are offered by Srivastava, John, Gosling, and Potter (2003). Indeed, evidence for dynamics in trait-like individual differences necessitates a rethinking of the logic presented in support of the supremacy of cohort effects in CTMA.

Second, there is ample evidence for the potent influence of period effects on behavior; indeed, this notion is fundamental to our understanding of social influence. Rudolph and Zacher (2017) review multiple studies of the influence of contemporaneous period effects on work-related behaviors (e.g., ranging from the influence of influenza epidemics, to the effects of day-light savings time "shifts," to the impact of college basketball tournaments on productivity). Moreover, research by Bianchi points to the influence of contemporaneous unemployment rates on differences in narcissism (Bianchi, 2014).

More recently, Eschleman, King, Mast, Ornellas, and Hunter (2017) demonstrated that variation in entitlement (a facet of narcissism) can be momentarily activated. This research suggests that generational differences are in part influenced by stereotype activation, and that merely thinking about generations may be enough to activate stereotype consistent responses. Finally, research that uses ecological momentary assessments to test theories of short-term intraindividual variability posits relationships between the experience of contemporaneous events and momentary dynamics in behaviors and attitudes (see Bolger & Laurenceau, 2013). To this end, recent research demonstrates that there is meaningful momentary variability in within-person levels of narcissism (Edershire, Woods, Sharpe, Crowe, Miller, & Wright, 2018; Walters & Horton, 2015).

Statistical Assumptions of CTMA

CTMA has appeal because it ties cohort membership directly to observed phenomena of interest (e.g., work values, see Twenge et al., 2010). Thus, CTMA represents a rather parsimonious representation of the inherent complexities of the age, period, and cohort problem (Glenn, 1976). Beyond the theoretically based assumptions reviewed previously, CTMA also makes the more statistically grounded assumption that cohort effects can be represented as changes over time in mean-levels of a given phenomenon. Indeed, the typical CTMA addresses a research question that follows something like, "Is there a demonstrable change in mean-levels of a given construct of interest over time?" Interestingly, the idea that mean-level changes are an appropriate metric to understand cohort effects has not been questioned in the CTMA literature. Regardless of whether or not this is theoretically appropriate, this assumption is testable, inasmuch as we can determine if

CTMA is capable of accurately capturing such mean-level changes. With this in mind, we now turn our attention to our Monte Carlo simulation, which addresses this point.

Monte Carlo Simulation Study

To date, there has not been a systematic effort to “un-pack” the boundaries of inferences that can be gleaned from CTMA. The advantage of using a Monte Carlo simulation framework for understanding such inferences is that certain parameters can be “fixed” in a theoretical population from which samples are drawn and subsequently analyzed. Specifically, this methodology allows us to specify *a priori* the degree of change that is present across time in the population, and then sample studies from this population. If CTMA is well geared for capturing such changes, then one would expect CTMA models to faithfully and reasonably reproduce this change parameter.

As suggested, the separation of period and cohort effects is an intractable problem *even in the population* (i.e., criticisms levied here are not attributable to non-representative sampling among individual studies included in the analysis). No amount of statistical, methodological, or empirical maneuvering can solve this problem. Therefore, logic must be applied to parse the influence of confounded period and cohort effects (see also Bell & Jones, 2013, 2014). It is important to note that the Monte Carlo simulations presented here cannot solve this confounding problem, but they can tell us how sensitive CTMA models are to accurately detecting the known-to-be confounded effects of period and cohort. That is to say, if we can momentarily suspend one of our primary critiques of CTMA, we can focus on whether or not such models actually represent reality as we have defined it (i.e., which happens to be a messy, confounded amalgamation of period and cohort influences).

Method

All simulations were done within the R environment for statistical computing (R Core Team, 2016); code to replicate these simulations can be found at: <https://osf.io/mak6y/>. A pre-print version of this work can be found here: <https://psyarxiv.com/exskp/>. Although simulations allow for a great deal of flexibility in parameterization, their presentation can be critiqued for focusing on narrow ranges of parameters (i.e., often out of necessity, as there are infinite combinations thereof). As such, we have also created a Shiny web application, which will both allow interested readers to directly replicate the results presented here, and to explore any number of additional combinations of parameters to further explore the limits of the CTMA framework (available here: https://cortrudolph.shinyapps.io/CTMA_

[Simulation](#)). That said, the setup and process of our Monte Carlo simulations were as follows:

Our Monte Carlo simulation is designed to mimic the process of conducting a CTMA. Thus, at the heart of our simulations is a mechanism designed to sample study-level means from an assumed population of studies. We began by specifying *a priori* population parameters from which to simulate samples representing individual studies to be subjected to a series of CTMAs. More specifically, we defined populations representing an outcome of interest. The nature of the outcome is irrelevant to the simulation; however, for the sake of this example, let us assume that it is ostensibly of enough import to warrant conducting a CTMA. We defined our populations such that each outcome had a mean, $\mu = 5$. Of note, the specific value of μ is irrelevant to the nature of these simulations; it only serves to scale the distributional properties of the population from which studies are simulated. Rather, the more important parameters being manipulated here are the number of “studies” sampled from each population (K), the size of each “sample” (N), the standard deviation in the population (σ), and the associated mechanism that generates confounded cohort-period effects (J_{PCE} : a variable that indexes the amount of change, in population standard deviation units, of the outcome in any given year). Considering each of these parameters, we adopted a $2 (K) \times 2 (N) \times 3 (\sigma) \times 11 (J_{PCE})$ fully-factorial design, which we next describe in more detail.

First, for each simulation condition, we randomly and without replacement generated either $K = 25$ or $K = 50$ and $N = 200$ or $N = 400$ “person” samples (i.e., each observation therein reflecting scores or standing on the aforementioned outcome of interest) from the *a priori* defined population to build a CTMA database. Consistent with the logic of many CTMAs, we assumed that the age of respondents was held more-or-less constant in the population to isolate confounded period and cohort effects. For each condition, each of the K samples in this database represents a single “study” of sample size N , representing the outcome phenomenon conducted in a single year, such that the first study, T , represents the first year, and every additional study ($T + 1 \dots [K - 1]$) represents studies conducted across either a 25- or 50-year timeframe.

We chose $K = 25$ studies as a starting point for our simulations, because a power analysis (i.e., construed as a two-tailed exact test for bivariate normal correlations; code to reproduce this analysis is available here: <https://osf.io/mak6y/>) suggested that 25 observations (i.e., representing years in this case) is the optimal sample size to detect a correlation of $r_{xy} = 0.53$ at $p < 0.05$ (two-tailed) with 80% power. We chose a correlation of this specific magnitude for this power analysis, because Twenge et al. (2008) report a bivariate relationship between time and NPI scores of $\beta_{xy} = 0.53$ (*n.b.*, as previously suggested, in the bivariate case, wherein, there is a single predictor entered into a regression model, $\beta_{xy} = r_{xy}$). To further explore the sensitivity of CTMA to variations in K , we also

explored conditions in which $K = 50$ (i.e., to effectively double the range of studies considered in these analyses). Regardless of condition, we represented each study as a single year, because the nesting of studies within year presents issues of non-independence of observations, a problem that has not, to our knowledge, been acknowledged by existing CTMAs.

Likewise, we chose to draw $N = 200$ “person” samples as a starting point, because the average sample size across the studies considered by Twenge et al. (2008; Table 1) was approximately $N = 200$ ($N = 193.82$). Again, to further explore the sensitivity of the CTMA model to N , we also explored conditions in which $N = 400$ (i.e., again, to effectively double the sample size considered in these analyses).

Regardless of condition, our decision to sample equally sized samples avoids having to differentially weight each study in our subsequent analyses (i.e., we assume that each study has equal weight, because each reflects an equally-sized sample, N , randomly drawn from the same population without replacement). As such, there is no advantage (i.e., with respect to mitigating bias) to differentially weighting each study in our simulation. We made the decision to draw equally-sized samples for two additional reasons: First, because samples of different sizes would receive different weights in the final model, this differential weighting could serve as a confounding factor to our ability to accurately isolate the effect of interest. Second, various CTMAs found in the literature use different weighting strategies (e.g., sample size weighting, see Twenge & Campbell, 2001; inverse variance weighting, Twenge 1997a). As suggested previously, even though these different weighting strategies tend to yield functionally equivalent results (Hunter & Schmidt, 2004), we still wanted to rule this out as a confounding factor in our study.

To introduce additional variability into our simulations, and to represent three different population standard deviation “conditions” that might be present for a given outcome, we specified separate populations with standard deviations of $\sigma = 0.50$, $\sigma = 1.00$, and $\sigma = 1.50$, respectively. We chose $\sigma = 0.50$ as a starting point, guided by the idea that one-half of a standard deviation is considered to be a notable amount of variability in the population (Cohen, 1988; Rosenthal & Rosnow, 1984). As before, to explore the sensitivity of the CTMA model to σ , we also explored conditions that doubled (i.e., $\sigma = 1.00$) and tripled ($\sigma = 1.50$) this value.

Finally, to model confounded period and cohort effects in the data that resulted from the aforementioned sampling procedures, we designed a period-cohort effect (PCE) generator to reflect confounded period-cohort effects assumed to operate in the population (henceforth, \bar{x}_{PCE}). We defined this \bar{x}_{PCE} generator as such:

$$\bar{x}_{PCE} = \bar{x}_{sample} + T_{1-K} \times (\sigma \times J_{PCE}) \tag{1}$$

where:

\bar{x}_{sample} is the arithmetic mean for any given sample, derived from our simulations, as described.

T_{1-K} is the previously noted time index (i.e., across K studies, representing each study’s “year” of data collection).

σ is the population standard deviation, defined in our population.

J_{PCE} is the “period-cohort effect multiplier,” a single value against which to increment the population standard deviation (σ) across time (T_{1-K}). As defined, J_{PCE} serves to index the amount of change, in population standard deviation units, of the outcome in any given year.

It is important to note that the typical “effects” reported in CTMAs vary to some extent. For example, Twenge et al. (2008) suggested a $d_{pseudo} = 0.33$ standard deviation change in narcissism over time. Taking this approximate value as an arbitrary “ballpark,” our simulations looked at a range of changes around this general value, but also considers more extreme values. We justified this based on the fact that the Twenge et al. (2008) paper has been cited nearly 930 times (i.e., according to a Google Scholar query on August 22, 2018). Thus, from the perspective of evidentiary value as reflected by citation counts, an effect of this magnitude seems to represent a “notable” finding (i.e., worthy of continued research; worthy of shaping policy; influencing discourse about this phenomenon). We also note that the effect sizes considered here are consistent with the range of average effect sizes reported in psychological research (see Szucs & Ioannidis, 2017)—although we make no assertion about the meaningfulness of the size of such an effect.

With this in mind, we considered 11 different conditions each representing a single J_{PCE} ranging from 0.001 to 0.050 in magnitude. We considered each of these 11 conditions across the two levels of K (i.e., 25, 50), the two levels of N (i.e., 200, 400), and the three levels of σ (i.e., 0.50, 1.00, 1.50). Table 1 summarizes these different combinations of conditions across each level of J_{PCE} considered. As defined in Formula (1), J_{PCE} represents an index of the change in the population level (i.e., scaled in terms of population standard deviation units) of the outcome in a given year. Accordingly, for each simulation, we can derive a pseudo-effect size estimate, δ_{pseudo} (i.e., true population effect size), calculated as follows:

$$\delta_{pseudo} = J_{PCE} \times K \times \sigma \tag{2}$$

which represents the degree of change across the K “years” of studies considered in our simulation. For example, $K = 25$, $\sigma = 0.50$, and $J_{PCE} = 0.001$ correspond to a population $\delta_{pseudo} = 0.0125$. Likewise, $K = 50$, $\sigma = 1.50$, and $J_{PCE} = 0.050$ correspond to a population $\delta_{pseudo} = 3.750$.

The next step was to compute CTMA models for \bar{x}_{PCE} across the K studies; this was done iteratively for each of the conditions. For each simulation and subsequent CTMA model, we noted the δ_{pseudo} and d_{pseudo} and computed a value reflecting the degree of “bias,” defined as:

Table 1. Summary of Monte Carlo simulations

Conditions	$K = 25$					$K = 50$				
	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias
$\sigma = 0.50$	0.001	0.013	0.105	0.004	8.397	0.001	0.025	0.050	0.001	2.001
$N = 200$	0.005	0.063	0.199	0.008	3.178	0.005	0.125	0.244	0.005	1.951
	0.010	0.125	0.316	0.013	2.526	0.010	0.250	0.486	0.010	1.945
	0.015	0.188	0.433	0.017	2.308	0.015	0.375	0.729	0.015	1.943
	0.020	0.250	0.550	0.022	2.200	0.020	0.500	0.971	0.019	1.942
	0.025	0.313	0.667	0.027	2.134	0.025	0.625	1.213	0.024	1.941
	0.030	0.375	0.784	0.031	2.091	0.030	0.750	1.455	0.029	1.941
	0.035	0.438	0.901	0.036	2.060	0.035	0.875	1.698	0.034	1.940
	0.040	0.500	1.018	0.041	2.036	0.040	1.000	1.940	0.039	1.940
	0.045	0.563	1.135	0.045	2.018	0.045	1.125	2.182	0.044	1.940
	0.050	0.625	1.252	0.050	2.004	0.050	1.250	2.425	0.048	1.940
	$K = 25$					$K = 50$				
Conditions	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias
$\sigma = 1.00$	0.001	0.025	0.105	0.004	4.198	0.001	0.050	0.050	0.001	1.000
$N = 200$	0.005	0.125	0.199	0.008	1.589	0.005	0.250	0.244	0.005	0.970
	0.010	0.250	0.316	0.013	1.263	0.010	0.500	0.486	0.010	0.972
	0.015	0.375	0.433	0.017	1.154	0.015	0.750	0.729	0.015	0.971
	0.020	0.500	0.550	0.022	1.100	0.020	1.000	0.971	0.019	0.971
	0.025	0.625	0.667	0.027	1.067	0.025	1.250	1.213	0.024	0.971
	0.030	0.750	0.784	0.031	1.045	0.030	1.500	1.455	0.029	0.970
	0.035	0.875	0.901	0.036	1.030	0.035	1.750	1.698	0.034	0.970
	0.040	1.000	1.018	0.041	1.018	0.040	2.000	1.940	0.039	0.970
	0.045	1.125	1.135	0.045	1.009	0.045	2.250	2.182	0.044	0.970
	0.050	1.250	1.252	0.050	1.002	0.050	2.500	2.425	0.048	0.970
	$K = 25$					$K = 50$				
Conditions	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias
$\sigma = 1.50$	0.001	0.038	0.105	0.004	2.799	0.001	0.075	0.050	0.001	0.667
$N = 200$	0.005	0.188	0.199	0.008	1.059	0.005	0.375	0.244	0.005	0.650
	0.010	0.375	0.316	0.013	0.842	0.010	0.750	0.486	0.010	0.648
	0.015	0.563	0.433	0.017	0.769	0.015	1.125	0.729	0.015	0.648
	0.020	0.750	0.550	0.022	0.733	0.020	1.500	0.971	0.019	0.647
	0.025	0.938	0.667	0.027	0.711	0.025	1.875	1.213	0.024	0.647
	0.030	1.125	0.784	0.031	0.697	0.030	2.250	1.455	0.029	0.647
	0.035	1.313	0.901	0.036	0.687	0.035	2.625	1.698	0.034	0.647
	0.040	1.500	1.018	0.041	0.679	0.040	3.000	1.940	0.039	0.647
	0.045	1.688	1.135	0.045	0.673	0.045	3.375	2.182	0.044	0.647
	0.050	1.875	1.252	0.050	0.668	0.050	3.750	2.425	0.048	0.647
	$K = 25$					$K = 50$				
Conditions	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias	J_{PCE}	δ_{pseudo}	d_{pseudo}	PCM	Bias
$\sigma = 0.50$	0.001	0.013	0.054	0.002	4.328	0.001	0.025	0.062	0.001	2.499
$N = 400$	0.005	0.063	0.149	0.006	2.384	0.005	0.125	0.258	0.005	2.066
	0.010	0.125	0.268	0.011	2.141	0.010	0.250	0.503	0.010	2.010
	0.015	0.188	0.386	0.015	2.060	0.015	0.375	0.748	0.015	1.994
	0.020	0.250	0.505	0.020	2.020	0.020	0.500	0.993	0.020	1.985
	0.025	0.313	0.624	0.025	1.995	0.025	0.625	1.237	0.025	1.980
	0.030	0.375	0.742	0.030	1.979	0.030	0.750	1.482	0.030	1.976
	0.035	0.438	0.861	0.034	1.968	0.035	0.875	1.727	0.035	1.974

Table 1. (continued)

Conditions	<i>K</i> = 25					<i>K</i> = 50				
	<i>J</i> _{PCE}	δ_{pseudo}	<i>d</i> _{pseudo}	PCM	Bias	<i>J</i> _{PCE}	δ_{pseudo}	<i>d</i> _{pseudo}	PCM	Bias
	0.040	0.500	0.979	0.039	1.959	0.040	1.000	1.972	0.039	1.972
	0.045	0.563	1.098	0.044	1.952	0.045	1.125	2.216	0.044	1.970
	0.050	0.625	1.217	0.049	1.947	0.050	1.250	2.461	0.049	1.969
	<i>K</i> = 25					<i>K</i> = 50				
Conditions	<i>J</i>_{PCE}	δ_{pseudo}	<i>d</i>_{pseudo}	PCM	Bias	<i>J</i>_{PCE}	δ_{pseudo}	<i>d</i>_{pseudo}	PCM	Bias
$\sigma = 1.00$	0.001	0.025	0.054	0.002	2.164	0.001	0.050	0.062	0.001	1.250
<i>N</i> = 400	0.005	0.125	0.149	0.006	1.192	0.005	0.250	0.258	0.005	1.033
	0.010	0.250	0.268	0.011	1.071	0.010	0.500	0.503	0.010	1.006
	0.015	0.375	0.386	0.015	1.030	0.015	0.750	0.748	0.015	0.997
	0.020	0.500	0.505	0.020	1.010	0.020	1.000	0.993	0.020	0.993
	0.025	0.625	0.624	0.025	0.998	0.025	1.250	1.237	0.025	0.990
	0.030	0.750	0.742	0.030	0.990	0.030	1.500	1.482	0.030	0.988
	0.035	0.875	0.861	0.034	0.984	0.035	1.750	1.727	0.035	0.987
	0.040	1.000	0.979	0.039	0.979	0.040	2.000	1.972	0.039	0.986
	0.045	1.125	1.098	0.044	0.976	0.045	2.250	2.216	0.044	0.985
	0.050	1.250	1.217	0.049	0.973	0.050	2.500	2.461	0.049	0.984
	<i>K</i> = 25					<i>K</i> = 50				
Conditions	<i>J</i>_{PCE}	δ_{pseudo}	<i>d</i>_{pseudo}	PCM	Bias	<i>J</i>_{PCE}	δ_{pseudo}	<i>d</i>_{pseudo}	PCM	Bias
$\sigma = 1.50$	0.001	0.038	0.054	0.002	1.443	0.001	0.075	0.062	0.001	0.833
<i>N</i> = 400	0.005	0.188	0.149	0.006	0.795	0.005	0.375	0.258	0.005	0.689
	0.010	0.375	0.268	0.011	0.714	0.010	0.750	0.503	0.010	0.671
	0.015	0.563	0.386	0.015	0.687	0.015	1.125	0.748	0.015	0.665
	0.020	0.750	0.505	0.020	0.673	0.020	1.500	0.993	0.020	0.662
	0.025	0.938	0.624	0.025	0.665	0.025	1.875	1.237	0.025	0.660
	0.030	1.125	0.742	0.030	0.660	0.030	2.250	1.482	0.030	0.659
	0.035	1.313	0.861	0.034	0.656	0.035	2.625	1.727	0.035	0.658
	0.040	1.500	0.979	0.039	0.653	0.040	3.000	1.972	0.039	0.657
	0.045	1.688	1.098	0.044	0.651	0.045	3.375	2.216	0.044	0.657
	0.050	1.875	1.217	0.049	0.649	0.050	3.750	2.461	0.049	0.656

*J*_{PCE}, period-cohort effect multiplier; δ_{pseudo} , population pseudo-effect size; *d*_{pseudo}, CTMA effect size; PCM, CTMA predicted period-cohort effect multiplier (i.e., J_{PCE}^{\wedge})

$$\text{bias} = d_{\text{pseudo}} / \delta_{\text{pseudo}} \tag{3}$$

Therefore, *bias* in this case indexes the extent to which the CTMA estimate (*d*_{pseudo}) misestimates the parameter that reflects true change in the population (δ_{pseudo}). Recall that because the parameter, δ_{pseudo} , is fixed in the population by virtue of the design of our simulations (i.e., \bar{x}_{PCE} ; see Formula (1)), any deviation—positive or negative—from this value reflects a misestimation of this parameter. For any given analysis, some deviation should be expected by random chance alone; however, systematic patterns of deviations may raise questions regarding the validity of conclusion about the true nature of the confounded period and cohort mechanism from the data.

For each simulation, we computed the CTMA following the procedures outlined by Twenge et al. (2008). We specified a sample size weighted least squares (WLS) regression model, regressing \bar{x}_{PCE} onto study year. Note that because our simulations only consider studies of the same sample size (*n*), the use of a WLS model here is equivalent to an ordinary least squares (OLS) model. Nonetheless, we use WLS here to allow for future research to more easily extend our simulations to consider variations in sample size across studies. To compute *d*_{pseudo}, we predicted values for the last (\hat{Y}_K) and first years (\hat{Y}_1) from the resulting regression equation and then divided the difference between these two predicted values by the average of the sample SDs across all *K* studies (see Twenge et al. 2008, p. 884):

$$d_{\text{pseudo}} = (\hat{Y}_K - \hat{Y}_1) / [(\sum SD) / K] \tag{4}$$

The resulting estimate indexes the predicted amount of change, in standard deviation units, across the $K = 25$ studies considered in each simulated CTMA.

Results

Table 1 reports the results of these simulations across the various conditions and Fig. 1 summarizes these results graphically. Under some conditions, CTMA efficiently estimates the population parameter. For example, when $K = 50$, $N = 200$, $\sigma = 1.00$, and $J_{\text{PCE}} = 0.001$, no bias is observed in the estimate (i.e., bias = 1.00, suggesting no difference between the estimate and the parameter), and when $K = 25$, $N = 400$, $\sigma = 1.00$, and $J_{\text{PCE}} = 0.025$, the observed bias = 0.998, suggesting the CTMA estimate only slightly underestimates the population parameter.

However, we temper this enthusiasm with two important observations. First, when both the population standard deviation and the simulated period-cohort effect are relatively small (i.e., less variability in the population and smaller changes in the population mean over time), the model gives estimates that are up to eight-times *higher* than the actual population parameter. Specifically, when $K = 25$, $N = 200$, $\sigma = 0.50$, and $J_{\text{PCE}} = 0.001$, the bias observed is 8.397 (see Fig. 2, panel A). Second, when both the population standard deviation and the simulated period-cohort effect are relatively high (i.e., corresponding to more variability in the population, and larger changes in the population over time), the model gives estimates that are over one-third *lower* than the actual population parameter. For example, when $K = 50$, $N = 200$, $\sigma = 1.50$, and

$J_{\text{PCE}} = 0.050$, bias = 0.647 (see Fig. 2, panel B). To visualize these differences, Fig. 2 presents slope plots (Tufté, 1983) representing values of the population parameter and the corresponding estimate across values of these two “worst case scenario” conditions (i.e., so-called, because these conditions represent the most extreme over- and under-estimates of the parameter, respectively).

To further understand how the parameters K , N , σ , and J_{PCE} affect the conclusions drawn from CTMA models, we specified a Type III ANOVA model, predicting the difference between the population parameter and that predicted from each CTMA model (i.e., difference = $\delta_{\text{pseudo}} - d_{\text{pseudo}}$). To interpret this difference, positive values indicate CTMA predictions that are greater than the parameter, whereas negative values indicate CTMA predictions that are less than the parameter. The variables K , N , σ , and J_{PCE} were entered into this model as independent fixed effects to ascertain their influence on this value. Table 2 summarizes the fit of this model; Table 3 summarizes the individual parameter estimates for these fixed effects. Data and code to reproduce these analyses are available here: <https://osf.io/mak6y/>.

Considering the results in Tables 2 and 3, only σ was a significant predictor of the observed difference ($p < 0.05$). To further understand this relationship, we estimated marginal means from this model and plotted them along with Bonferroni-corrected confidence intervals (i.e., 99.983%, reflecting 95% confidence intervals adjusted for the familywise error rate associated with three pairwise comparisons; see Fig. 3). It is clear that the CTMA is most likely to overestimate the parameter when σ is “low” (i.e., $\sigma = 0.50$, $M_{\text{Difference}} = 0.468$) and underestimate the parameter when σ is “high” (i.e., $\sigma = 1.50$, $M_{\text{Difference}} = -0.473$). The best chance for

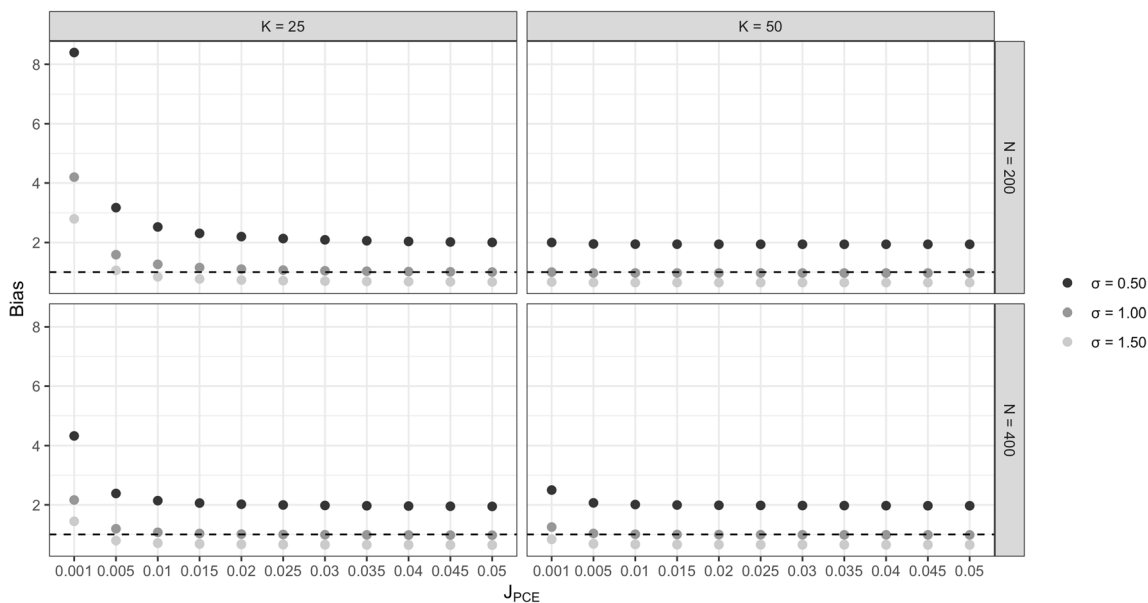
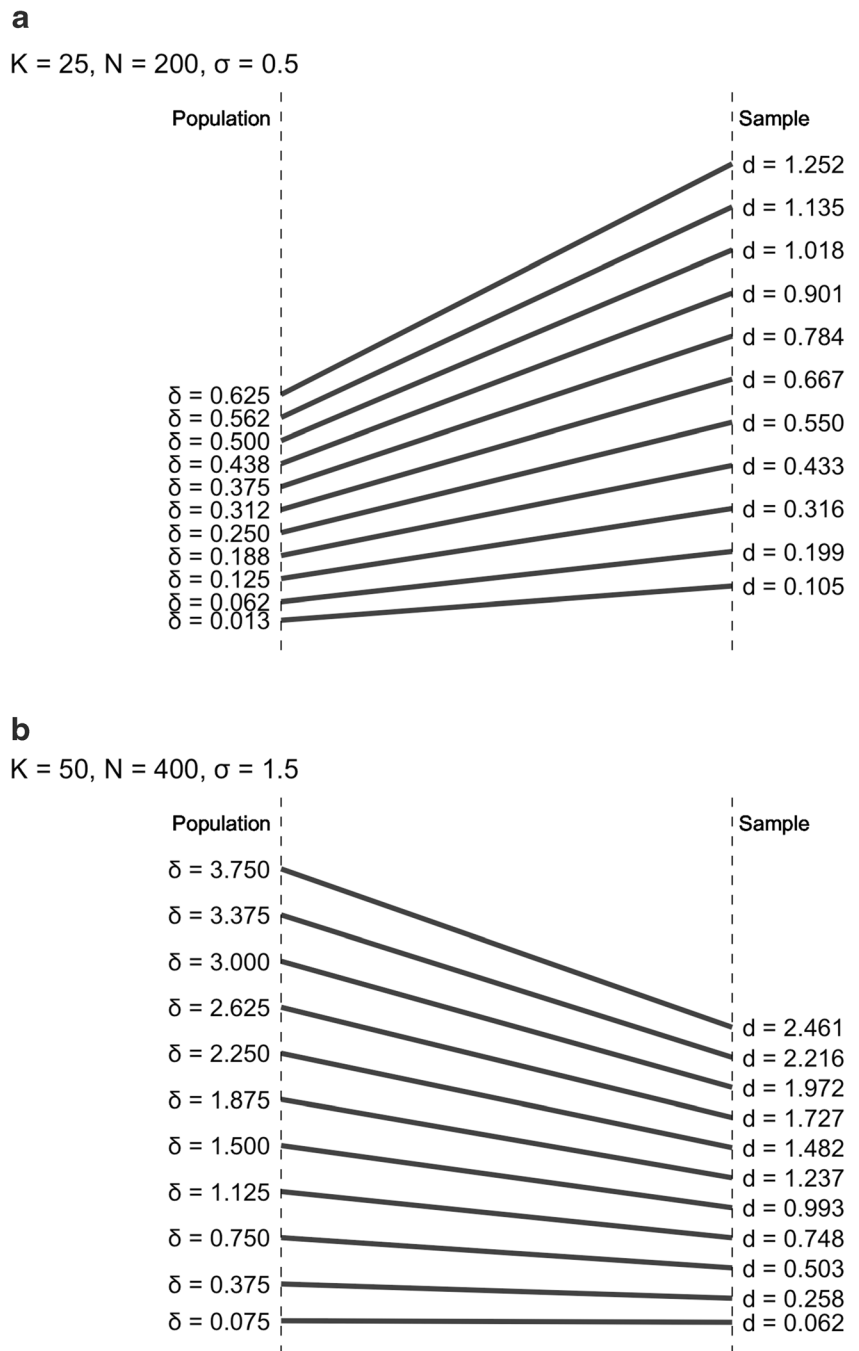


Fig. 1. Results of Monte Carlo simulations. Dashed horizontal line indicates bias = 1.00 (i.e., no bias)

Fig. 2. Slope plots for “worst case scenario” conditions



CTMA to correctly estimate the parameter occurs when σ is “moderate” (i.e., $\sigma = 1.00, M_{\text{Difference}} = -0.002$).

Moreover, mirroring the observations outlined previously, the confidence interval for this estimate includes zero, suggesting this condition is most likely to be unbiased with respect to estimating the parameter. Despite this, given the narrow range of cases for which unbiased estimates could be observed, the results presented here support the conclusion that the use of CTMA should be questioned (Donnellan et al., 2009; Trzesniewski et al., 2008). Indeed, except for a relatively narrow range of cases, our Monte Carlo simulation

suggests that CTMA is likely to offer quite different estimates of known population parameters.

Discussion

CTMA has become an increasingly popular method to examine purported generational cohort effects—that is, average differences in attitudes, values, or behaviors between groups of people based on ranges of birth cohorts. Increasingly, conclusions from CTMAs are taken as evidence for generational differences, the

Table 2. Summary of ANOVA model effects

Parameter	SS	df	F	P	η^2	η^2_{partial}	F_{Cohens}
(Intercept)	0.029	1	0.321	0.572	0.001	0.003	0.052
K	0.067	1	0.736	0.393	0.002	0.006	0.079
N	0.003	1	0.033	0.855	0.000	0.000	0.017
σ	19.477	2	107.049	0.000	0.643	0.647	1.353
J_{PCE}	0.06	10	0.066	1.000	0.002	0.006	0.075
Residuals	10.644	117					

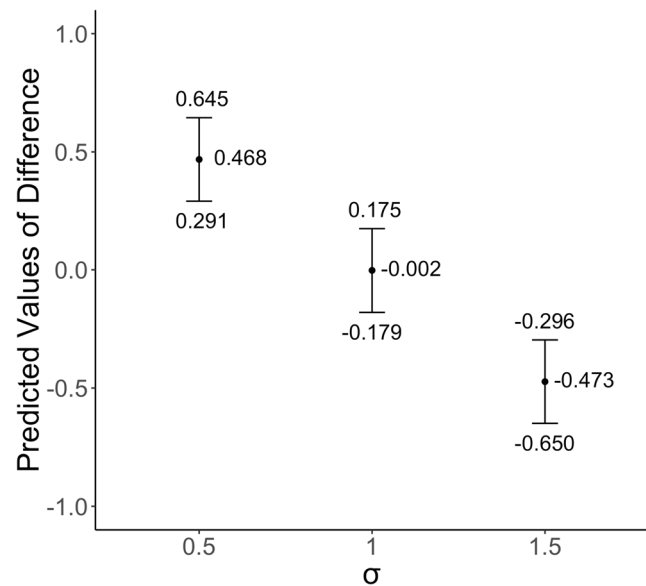
$N = 132$; $R^2 = 0.648$; $R^2_{\text{adjusted}} = 0.606$; Omnibus model: $F_{(14, 117)} = 15.394$, $p < 0.001$

implications of which have been widely discussed in and extended to the IO/OB/HR literatures. CTMA results are appealing, and many have adopted their interpretation that important differences between generations exist and need to be actively addressed. Here, we took a critical perspective on CTMA by scrutinizing its theoretical and methodological underpinnings.

We argued that even if meaningful generational existed, CTMA would not be able to tease apart period and cohort effects. Recent work that rests upon CTMA evidence has acknowledged this fact but also argued that this differentiation is irrelevant (e.g., "... practically speaking, it might not always matter much whether a change is due to generation or time period, because both indicate cultural change;" Twenge, 2017, Appendix A, p. 4). We would argue that whether such effects are due to period or

Table 3. Summary of ANOVA model parameters

	B	SE_B	95% C.I.	
			Lower	Upper
(Intercept)	0.06	0.10	-0.14	0.26
$K = 50$	-0.05	0.05	-0.15	0.06
$N = 400$	-0.01	0.05	-0.11	0.09
$\sigma =$				
0.50	0.47	0.06	0.34	0.60
1.50	-0.47	0.06	-0.60	-0.34
$J_{\text{PCE}} =$				
0.005	-0.01	0.12	-0.25	0.24
0.010	-0.01	0.12	-0.26	0.23
0.015	-0.02	0.12	-0.26	0.23
0.020	-0.03	0.12	-0.27	0.22
0.025	-0.03	0.12	-0.28	0.21
0.030	-0.04	0.12	-0.28	0.20
0.035	-0.05	0.12	-0.29	0.20
0.040	-0.05	0.12	-0.30	0.19
0.045	-0.06	0.12	-0.30	0.18
0.050	-0.07	0.12	-0.31	0.18
N	132			
$R^2/R^2_{\text{adjusted}}$	0.648/0.606			

**Fig. 3.** Estimated marginal means from ANOVA model

cohort influences does matter for the enactment of policies, within organizations and beyond. For example, organizations that have been advised to cater their human resources policies and practices to certain generations (e.g., Benson et al., 2018) might well be wasting valuable resources if it turns out that what younger employees want (i.e., an assumed cohort effect) is also in fact what all employees want (i.e., a period effect).

Beyond this, the notion of generational differences suffers from a number of conceptual problems that have been noted elsewhere, such as "fuzzy" boundaries (i.e., arbitrary birth year cut-offs) and conclusions biased by the ecological nature of the relationships examined (i.e., using group-level results to draw conclusions about individual-level relationships; Rudolph & Zacher, 2017). Furthermore, proponents have argued that CTMA controls for age effects when studying confounded period and cohort effects. However, this argument neglects to recognize that even in samples of college students (i.e., the most common samples used in CTMA, e.g., Twenge, 1997a, 1997b; Twenge et al., 2008), differences that emerge between 18- through 22-year-olds (e.g., via developmental processes or socialization) result in a certain degree of age-related variability. This age-related variability is further confounded with both period and cohort effects. For example, Robins, Fraley, Roberts, and Trzesniewski (2001) show that personality changes across the course of a standard four-year undergraduate degree program, and Roberts, Caspi, and Moffitt (2001) demonstrate similar patterns of variability between ages 18 and 26.

We empirically examined CTMA assumptions about the estimation of confounded period and cohort effects using a Monte Carlo study. Our simulations show that confounded period and cohort effects are likely to be misestimated, in some cases systematically overestimated or underestimated by CTMA. Even if CTMA gave reasonably accurate estimates of the parameter in

question, the estimate would still reflect confounded period and cohort effects making its utility questionable for both practical and theoretical purposes. Overall, then, the most important implication of the present paper is that the practice of conducting CTMA should be viewed with a high-degree of skepticism.

Revisiting three exemplar studies cited in the Introduction, one can see the varying impact that our findings have on research that has adopted the CTMA methodology. For example, Twenge and Campbell (2001) utilized CTMA to examine purported generational changes in self-esteem. This study concluded that, depending on the time period, self-esteem levels have changed over time. Given our results, there are at least two distinct problems with this conclusion that are common to many studies that use CTMA. First, as noted previously, CTMA typically confounds period and cohort effects with one-another when defining generations, doing nothing to separate these two sources of variance. Second, as our results show, there is very likely some degree of misestimation of the effect size attributable to this confounded period-cohort effect. In the case of Twenge and Campbell (2001), both of these issues are present, raising concerns about the validity of the conclusions drawn from this work. Moreover, the same critique could be levied against any CTMA that possesses these two qualities.

Similarly, Yang, et al. (2014) investigated changes in anxiety among Chinese military personnel using CTMA. This study concluded that both state and trait anxiety levels have increased over time in this population. Unlike other CTMAs (e.g., Twenge & Campbell, 2001; Twenge, et al. 2008), this study does not confound period and cohort effects, because it did not attempt to examine cohort effects through the process of isolating age (e.g., by considering only samples representing narrow age ranges). That said, Yang et al. (2014) still suffers from issues related to the misestimation of effect sizes due to the biased nature of such estimates, as highlighted by our study. Indeed, both state and trait anxiety show large population standard deviations (i.e., ± 8 points on an 80-point scale) and effect sizes (i.e., 0.88 and 0.63, respectively). Given our findings, this suggests that the estimates offered by Yang et al. (2014) may be misestimating this degree of change, again calling the validity of these conclusions into question.

A final example comes from the CTMA presented by Wegman et al. (2018), which examined whether workers' perceptions of job characteristics have changed between 1975 and 2011. This study concluded that workers perceive greater levels of skill variety, autonomy, and interdependence over time. The CTMA technique used in this study is unique, in that although it did examine cohort effects, it avoided confounding cohort and period by not conceptualizing cohorts as defined by age or period. Specifically, cohorts were operationalized as groups defined by high, middle, or low occupational complexity. Although this study may seem like a model case in which CTMA could be used appropriately, it is important to keep in mind that this study also falls victim to the notable limitations

underlying the misestimation of effect sizes that our study highlights. As with the previous two examples, the validity of these conclusions should be scrutinized accordingly.

Beyond the implications of our findings for research that has specifically used CTMA, a great deal of research has investigated work-related phenomena that might be impacted by age, period, and/or cohort effects. In particular, such research concerns how these effects relate to assumed generational differences in work outcomes. To this end, a recent review of the IO/OB/HR literatures on generational differences (Costanza, Finkelstein, Imose, & Ravid, *in press*) identified a set of common inferences that are drawn from such research, and offered a critical review of the appropriateness of such inferences. Extending the findings of the present effort to this review, and the inferences it identified sheds some light on the implications of our results to the broader domain of research on generational differences in work outcomes.

For example, one of the common inferences identified by Costanza et al. (*in press*) is that organizations should offer customized human resources (HR) policies and practices to account for generational differences. In some cases, this inference is supported by evidence drawn, often indirectly, from CTMAs (e.g., Lub et al., 2012; Ng & Johnson, & Burke, 2015; Stone & Deadrick, 2015). Given the limitations of the CTMA methodology highlighted here, known issues associated with other means of modeling age-period-cohort effects (Costanza et al., 2017; Schaie, 1965), and broader critiques of the generations concept found in the literature (e.g., Rudolph & Zacher, 2017), the reality is that there is no credible empirical evidence for designing and implementing customized, generationally-differentiated HR policies and practices. Furthermore, beyond simply wasting time and other resources, it is important to consider the potential legal problems that are associated with the provision of such differentiated policies. Indeed, organizations could easily end up offering HR policies and practices that employees do not need (or want), while simultaneously increasing their risk of litigation stemming from age-based differential treatment.

Given these issues, the question remains: "What is the alternative?" The search for a method to empirically separate age, period, and cohort effects has long eluded researchers (Schaie, 1965). As suggested by Glenn (2005), "The continued search for a statistical technique that can be mechanically applied always to correctly estimate the effects is one of the most bizarre instances in the history of science of repeated attempts to do the logically impossible" (p. 6). CTMA proponents have gone to great lengths to promote the influence of cohort effects, while downplaying the influence of period effects in such models (e.g., Gentile et al., 2015; Twenge et al., 2008). Such self-protective behaviors by CTMA proponents make sense, because if period effects were *really* the cause of variability in attitudes, values, and behavior, the entire idea of "generations" falls apart, along with the findings of nearly every CTMA ever produced.

At the risk of being glib, the “generations industry” (i.e., authors of popular press books, management consulting firms, generations “gurus”) has a vested interest in keeping afloat the idea that cohorts exhibit an important influence on people’s attitudes, values, behaviors, and patterns of thinking. Despite the evidence offered here, we suspect that the existence of generational differences will continue to be the subject of debate. However, based on our analyses and results, we are certain that CTMA results should not be the evidence upon which the pro-generational position is argued. There are a number of tangible risks associated with the continued use of CTMA results to develop theory, applications, and policies, including making poor decisions, wasting resources, reinforcing stereotypes, and ignoring/misattributing underlying causes of differences between people and groups.

Limitations and Directions for Future Research

Given the intractable nature of the confounding of period and cohort effects in CTMA, one might question whether our Monte Carlo simulations are even necessary to demonstrate this point. To this, let us suspend our arguments and assume that there is still value in understanding confounded period-cohort effects (i.e., as proponents of CTMA have argued). Our simulations show that the extent to which CTMA faithfully represents these known-to-be confounded parameters is still questionable in many cases. Thus, even with the argument that CTMA has never claimed to be a panacea for such confounded effects, it still does a sub-optimal job of estimating the parameters that it purports to.

Although our simulation results offer some important insights into the operation of CTMA, there are some limitations to note. As Bandalos and Gagne (2012) state, “Even the most elegantly designed [simulation] study may not be informative if the conditions included are not relevant to the type of data one typically encounters in practice” (p. 96). We purposefully chose parameters we believe were reasonable, inasmuch they are likely to be observed in primary research. That said, we also recognize that there are an unlimited number of possible conditions against which these models could be considered.

What we present here represents but a few plausible circumstances that one might expect to observe when conducting a CTMA and is not intended to be a comprehensive accounting of all possible derivations thereof. We invite interested readers to consider different combinations of such parameters using our Shiny web-application (linked above). Future research should also consider additional parameters, for example how assumptions about population distributions (i.e., we assumed normally distributed outcomes) and nonlinearity (i.e., we assumed a positive linear period-cohort effect) affect these conclusions.

Additionally, we were not able to address all potential limitations of CTMA empirically. Our argument that CTMA

unduly neglects period effects and overemphasizes cohort effects can only be a theoretical one. For instance, researchers have argued that there are numerous studies that provide evidence for influences of contemporaneous period effects on behavior (Rudolph & Zacher, 2017). Further theorizing and research on period effects is needed to make an even stronger case against the practice of downplaying such effects on attitudes, values, and behavior in CTMA.

With our Monte Carlo simulation, we add to the growing body of research that has pointed out limitations of analytic approaches used to study cohort effects (Costanza et al., 2017; Rudolph & Zacher, 2017; Trzesniewski & Donnellan, 2010). Although the systematic overestimation of cohort effects has not been demonstrated in previous research, psychological scientists have empirically shown that CTMA results, which are based on ecological relationships, cannot be replicated at the individual level (Trzesniewski et al., 2008). Moreover, Trzesniewski et al. (2008) have noted that convenience samples (e.g., *ad-hoc* samples of college students) should not be used to draw inferences about generational differences, because the findings are likely to be biased due to sampling error. However, even if researchers based CTMA on large, randomly sampled representative samples, our simulation study suggests that confounded period and cohort effects will be overestimated in many cases, and in particular when the mechanism leading to such effects is small.

One limitation of our Monte Carlo simulation is that it was specifically designed for the purposes of imitating the process of conducting CTMA. That is, we designed our simulations to sample study-level means from an assumed population of studies. Thus, our simulations are limited in their ability to concurrently test other analytic models for parsing age, period, and cohort effects in the way that Costanza et al. (2017) have recently done with primary study data. Given that the goal of the present study was to unpack the workings of CTMA, this decision is appropriate here. However, we call for future studies to modify our simulation framework to allow for the sampling of individual-level data from an assumed population of people. Such simulations would serve to extend this work and that of Costanza et al. (2017) by allowing for the comparison of these analytic models against one another given known parameters representing age, period, and cohort effects.

Another limitation of CTMA is the assumption that, by holding age more-or-less constant (e.g., assuming college students are all approximately the same age), people are particularly susceptible to historical influences at younger ages (i.e., the so-called “crystallization and ratification argument,” see Rudolph & Zacher, 2017). Lifespan psychologists have similarly argued that history-graded influences are particularly strong between the ages of 10 and 40 years (Baltes, Reese, & Lipsitt, 1980). In contrast, age-graded normative influences (e.g., physical maturation and decline) are assumed to be most influential before the age of 20 and after the age of 50 years.

Finally, non-normative (or idiosyncratic) influences on development are very rare (e.g., accidents, winning the lottery) and their relative strength of influence is thought to increase linearly across the lifespan (Baltes et al., 1980). While the existence of such lifespan profiles of relative influences of age-graded, history-graded, and non-normative influences on development makes intuitive sense, their validity has so far not been tested empirically. Doing so would prove to be challenging, as researchers would have to follow participants across their entire lifespan and assess the relative impact many different person-related and environmental influences.

Our hope, at this point, is that readers will recognize the problems inherent in CTMA. We are left then to ponder whether or not there are potential alternatives to the use of CTMA for studying generational influences on attitudes, values, and behavior. The short answer to this is “no.” The commonly used alternatives, cross-sectional and simple longitudinal research designs—approaches often employed in research conducted in organizational settings—are not useful for studying generational differences because they are not able to disentangle age, period, and cohort effects. Hence, our concern is that researchers and organizations may turn to CTMA to assess generational differences among their members, under the misguided assumption that it circumvents such issues.

More advanced designs, such as cohort-sequential designs, sample participants from different birth cohorts repeatedly over a certain period of time (Hofer & Sliwinski, 2006). The advantages of such designs are that age and cohort effects are not confounded and participants can be compared to their own baseline levels of a variable at the intraindividual level. Such designs are not advantageous to the study of generations, however, because period effects are still confounded with cohort effects. Finally, time-sequential designs involve two or more cross-sectional studies conducted at different measurement periods. In contrast to cohort-sequential designs, time-sequential designs allow for unconfounded age and period effects, but not cohort effects (see Rudolph & Zacher, 2017, for a review of various developmental research designs).

At this point, readers may also wonder whether there is a tenable answer to the question, “What analytic framework is appropriate for studying generational differences?” Unfortunately, there is no clear answer to this question, or at least one that can reasonably address this satisfies the strict condition of non-confounded age, period, and/or cohort effects. For example, even cross-classified hierarchical linear models (CCHLM), which have previously been offered as a solution to period-cohort confounding, are very limited in their capacity to disentangle these effects (i.e., they are only appropriate when age and period are *not* used to define cohort membership and variables are assigned to different levels of analysis arbitrarily). Researchers have gone so far as to characterize CCHLM as yet another “futile quest” to solve this intractable problem (see Bell & Jones, 2013, 2014). Indeed,

we know of no method that can efficiently and independently parse age, period, and/or cohort effects from one another. Given the results of our Monte Carlo simulations, consider that even in cases where the CTMA model seems to perform well (i.e., with respect to estimating a given parameter), the parameter that is being estimated is still a confounded period-cohort effect, which does not unequivocally inform an understanding of the influence of either period *or* cohort effects.

Indeed, as noted by Glenn in the preface to his treatise on cohort analysis, “... except under conditions that hardly ever exist, a definitive separation of age, period, and cohort effects is not just difficult, but impossible” (2005, p. vii). The only thing that we can be certain of from such models is that age is likely not a contributing factor to the variance being explained. Although, the general practice of sampling individuals from age ranges (e.g., “college students”) rather than those of discrete ages (e.g., “18-year-olds”) greatly complicates this conclusion.

Given these limitations, we currently do not have methodological and statistical tools available that allow studying generational differences in an unambiguous way. As such, some have even gone so far as to suggest that there has never actually been a study of generational differences (Rudolph & Zacher, 2018) and has led scholars to call for a moratorium to be placed upon research on such differences (e.g., Rudolph & Zacher, 2017; Rudolph et al., 2018).

Conclusion

We explored the background, history, and assumptions of CTMA in this paper. We argued that CTMA is not able to clearly disentangle age, period, and cohort effects. Moreover, the results of a Monte Carlo study showed that, even if CTMA was able to disentangle these developmental influences, it is likely to systematically overestimate confounded period-cohort effects. Thus, we recommend that a great deal of caution should be exercised in applying CTMA and interpreting its findings to suggest that generational differences exist.

References

- Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development, 11*, 145–171. <https://doi.org/10.1159/000270604>.
- Baltes, P. B., Reese, H. W., & Lipsitt, L. P. (1980). Life-span developmental psychology. *Annual Review of Psychology, 31*(1), 65–110. <https://doi.org/10.1146/annurev.ps.31.020180.000433>.
- Bandalos, D. L., & Gagné, P. (2012). Simulation methods in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural*

- equation modeling* (pp. 92–108). New York, NY, US: Guilford Press.
- Bell, A., & Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social Science & Medicine*, *93*, 163–165. <https://doi.org/10.1016/j.socscimed.2013.04.029>.
- Bell, A., & Jones, K. (2014). Another 'futile quest'? A simulation study of Yang and Land's hierarchical age-period-cohort model. *Demographic Research*, *30*, 333–360. <https://doi.org/10.4054/DemRes.2013.30.11>.
- Benson, J., Brown, M., Glennie, M., O'Donnell, M., & O'Keefe, P. (2018). The generational "exchange" rate: How generations convert career development satisfaction into organisational commitment or neglect of work. *Human Resource Management Journal*, *28*(4), 524–539. <https://doi.org/10.1111/1748-8583.12198>.
- Bianchi, E. C. (2014). Entering adulthood in a recession tempers later narcissism. *Psychological Science*, *25*, 1429–1437. <https://doi.org/10.1177/0956797614532818>.
- Blyth, C. R. (1972). On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, *67*(338), 364–366. <https://doi.org/10.1080/01621459.1972.10482387>.
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.
- Bubany, S. T., & Hansen, J. I. C. (2011). Birth cohort change in the vocational interests of female and male college students. *Journal of Vocational Behavior*, *78*(1), 59–67. <https://doi.org/10.1016/j.jvb.2010.08.002>.
- Campbell, S. M., Twenge, J. M., & Campbell, W. K. (2017). Fuzzy but useful constructs: Making sense of the differences between generations. *Work, Aging and Retirement*, *3*(2), 130–139. <https://doi.org/10.1093/workar/wax001>.
- Clark, D. M. T., Loxton, N. J., & Tobin, S. J. (2015). Declining loneliness over time: Evidence from American colleges and high schools. *Personality and Social Psychology Bulletin*, *41*(1), 78–89. <https://doi.org/10.1177/0146167214557007>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Costanza, D. P., Darrow, J. B., Yost, A. B., & Severt, J. B. (2017). A review of analytical methods used to study generational differences: Strengths and limitations. *Work, Aging and Retirement*, *3*(2), 149–165. <https://doi.org/10.1093/workar/wax002>.
- Costanza, D. P., Darrow, J. B., Fraser, R. L., Severt, J. B., & Gade, P. A. (2012). Generational differences in work-related variables: A meta-analysis. *Journal of Business and Psychology*, *27*(4), 375–394. <https://doi.org/10.1007/s10869-012-9259-4>.
- Costanza, D. P., & Finkelstein, L. M. (2015). Generationally-based differences in the workplace: Is there a there there? *Industrial and Organizational Psychology: Perspectives on Sciences and Practice*, *8*(3), 308–323. <https://doi.org/10.1017/iop.2015.15>.
- Costanza, D. P., Finkelstein, L. M., Imose, R. A., & Ravid, D. M. in press Inappropriate inferences from generational research. In B. Hoffman, M. Shoss, & L. Wegman (Eds.), *The Cambridge handbook of the changing nature of work*. Cambridge, U.K.: Cambridge University Press.
- Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2009). An emerging epidemic of narcissism or much ado about nothing? *Journal of Research in Personality*, *43*(3), 498–501. <https://doi.org/10.1016/j.jrp.2008.12.010>.
- Edershile, E. A., Woods, W. C., Sharpe, B. M., Crowe, M. L., Miller, J., & Wright, A. G. (2018, July 12). *A day in the life of Narcissus: Measuring narcissistic grandiosity and vulnerability in daily life*. <https://doi.org/10.31234/osf.io/jpqst>.
- Elder Jr., G. H. (1974). *Children of the great depression: Social change in life experiences*. Chicago, IL: University of Chicago Press.
- Elder Jr., G. H., & Liker, J. K. (1982). Hard times in women's lives: Historical in influences across forty years. *American Journal of Sociology*, *88*, 241–269. <https://doi.org/10.1086/227670>.
- Eschleman, K. J., King, M., Mast, D., Ornellas, R., & Hunter, D. (2017). The effects of stereotype activation on generational differences. *Work, Aging and Retirement*, *3*(2), 200–208. <https://doi.org/10.1093/workar/waw032>.
- Gentile, B., Wood, L. A., Twenge, J. M., Hoffman, B. J., & Campbell, W. K. (2015). The problem of generational change: Why cross-sectional designs are inadequate for investigating generational differences. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical myths and methodological urban legends*. New York: Routledge.
- Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*(2), 309–320. <https://doi.org/10.1037/h0034436>.
- Gerstorff, D., Ram, N., Hoppmann, C., Willis, S. L., & Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology*, *47*, 1026–1041. <https://doi.org/10.1037/a0023426>.
- Glenn, N. D. (1976). Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. *American Sociological Review*, *41*(5), 900–904. <https://doi.org/10.2307/2094738>.
- Glenn, N. D. (2005). *Cohort analysis* (2nd ed.). London: Sage.
- Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science*, *5*(1), 33–36. <https://doi.org/10.1111/j.1467-9280.1994.tb00610.x>.
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 14–50). Baltimore, MD: Johns Hopkins University Press.
- Hofer, S. M., & Sliwinski, M. J. (2006). Design and analysis of longitudinal studies on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 15–37). Amsterdam: Elsevier.
- Huang, J. (2018). Changes of job burnout in Chinese nurses over 2004–2013: Cross-temporal meta-analysis. *Current Psychology*, *37*(3), 583–590. <https://doi.org/10.1007/s12144-016-9540-1>.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Karaszia, B. T., Tylka, T. L., & Murnen, S. K. (2017). Is body dissatisfaction changing across time? A cross-temporal meta-analysis. *Psychological Bulletin*, *143*(3), 293–320. <https://doi.org/10.1037/bul0000081>.
- Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, *19*, 195–229. <https://doi.org/10.1177/109442810033001>.
- Kosloski, K. (1986). Isolating age, period, and cohort effects in developmental research: A critical review. *Research on Aging*, *8*(4), 460–479. <https://doi.org/10.1177/0164027586008004002>.
- Labouvie-Vief, G., Hakim-Larson, J., DeVoe, M., & Schoeberlein, S. (1989). Emotions and self-regulation: A life span view. *Human Development*, *32*(5), 279–299. <https://doi.org/10.1159/000276480>.
- Lub, X., Nije Bijvank, M., Bal, P. M., Blomme, R., & Schalk, R. (2012). Different or alike? Exploring the psychological contract and commitment of different generations of hospitality workers. *International Journal of Contemporary Hospitality Management*, *24*(4), 553–573. <https://doi.org/10.1108/09596111211226824>.
- Lyons, S. T., & Schweitzer, L. (2017). A qualitative exploration of generational identity: Making sense of young and old in the context of today's workplace. *Work, Aging and Retirement*, *3*(2), 209–224. <https://doi.org/10.1093/workar/waw024>.
- Mackenzie, C., Erickson, J., Deane, F., & Wright, M. (2014). Changes in attitudes toward seeking mental health services: A 40-year cross-

- temporal meta-analysis. *Clinical Psychology Review*, 34(2), 99–106. <https://doi.org/10.1016/j.cpr.2013.12.001>.
- Malahy, L. W., Rubinlicht, M. A., & Kaiser, C. R. (2009). Justifying inequality: A cross-temporal investigation of U.S. income disparities and just-world beliefs from 1973 to 2006. *Social Justice Research*, 22, 369–383. <https://doi.org/10.1007/s11211-009-0103-6>.
- Ng, E. S., Johnson, J. M., & Burke, R. J. (2015). Millennials: Who are they, how are they different, and why should we care. In R. J. Burke, C. Cooper, & A.-S. Antoniou (Eds.), *The multi-generational and aging workforce challenges and opportunities*. Cheltenham, U.K.: Edward Elgar Publishing.
- Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology*, 78(4), 569–582. <https://doi.org/10.1037/0021-9010.78.4.569>.
- Ostroff, C., & Harrison, D. A. (1999). Meta-analysis, level of analysis, and best estimates of population correlations: Cautions for interpreting meta-analytical results in organizational behavior. *Journal of Applied Psychology*, 84(2), 260–270. <https://doi.org/10.1037/0021-9010.84.2.260>.
- Palmore, E. (1978). When can age, period, and cohort be separated? *Social Forces*, 57(1), 282–295. <https://doi.org/10.2307/2577639>.
- Perry, E. L., Golom, F. D., Catenacci, L., Ingraham, M. E., Covais, E. M., & Molina, J. J. (2017). Talkin’ ‘bout your generation: The impact of applicant age and generation on hiring-related perceptions and outcomes. *Work, Aging and Retirement*, 3(2), 186–199. <https://doi.org/10.1093/workar/waw029>.
- Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Pervasiveness of the IQ rise: A cross-temporal meta-analysis. *PLoS ONE*, 5(12), e14406. <https://doi.org/10.1371/journal.pone.0014406>.
- Core Team, R. (2016). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna: Austria. URL <https://www.R-project.org/>.
- Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology*, 81(4), 670–683. <https://doi.org/10.1037/0022-3514.81.4.670>.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. <https://doi.org/10.1037/0033-2909.132.1.1>.
- Roberts, B. W., & Wood, D. (2006). Personality development in the context of the neo-socioanalytic model of personality. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 11–39). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, 69(4), 617–640. <https://doi.org/10.1037/0033-2909.132.1.1>.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357. <https://doi.org/10.2307/2087176>.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59–82. <https://doi.org/10.1146/annurev.psych.52.1.59>.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.
- Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in research: A correlational approach*. New York: Cambridge University Press.
- Rauvola, R. S., Rudolph, C. W., & Zacher, H. (2018). Generationalism: Problems and implications. *Organizational Dynamics*. <https://doi.org/10.1016/j.orgdyn.2018.05.006>.
- Rudolph, C. W., Rauvola, R. S., & Zacher, H. (2018). Leadership and generations at work: A critical review. *The Leadership Quarterly*, 29(1), 44–57. <https://doi.org/10.1016/j.leaqua.2017.09.004>.
- Rudolph, C. W. (2016). Lifespan developmental perspectives on working: A literature review of motivational theories. *Work, Aging and Retirement*, 2(2), 130–158. <https://doi.org/10.1093/workar/waw012>.
- Rudolph, C. W., & Baltes, B. B. (2016). Age and health jointly moderate the influence of flexible work arrangements on work engagement: Evidence from two empirical studies. *Journal of Occupational Health Psychology*, 22(1), 40–58. <https://doi.org/10.1037/a0040147>.
- Rudolph, C. W., & Zacher, H. (2015). Intergenerational perceptions and conflicts in multi-age and multigenerational work environments. In L. Finkelstein, D. Truxillo, F. Fraccaroli, & R. Kanfer (Eds.), *SIOP organizational frontier series—Facing the challenges of a multi-age workforce: A use inspired approach* (pp. 253–282). New York, NY: Psychology Press.
- Rudolph, C. W., & Zacher, H. (2017). Considering generations from a lifespan developmental perspective. *Work, Aging and Retirement*, 3(2), 113–129. <https://doi.org/10.1093/workar/waw019>.
- Rudolph, C. W., & Zacher, H. (2018). The kids are alright: Taking stock of generational differences at work. *The Industrial-Organizational Psychologist*, 55(3), 1–7.
- Schaie, K. W. (1986). Beyond calendar definitions of age, time, and cohort: The general developmental model revisited. *Developmental Review*, 6(3), 252–277. [https://doi.org/10.1016/0273-2297\(86\)90014-6](https://doi.org/10.1016/0273-2297(86)90014-6).
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, 64(2), 92.
- Schaie, K. W. (2013). *Developmental influences on adult intelligence: A Seattle longitudinal study* (2nd ed.). New York, NY: Oxford University Press.
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84(5), 1041–1053. <https://doi.org/10.1037/0022-3514.84.5.1041>.
- Stone, D. L., & Deadrick, D. L. (2015). Challenges and opportunities affecting the future of human resource management. *Human Resource Management Review*, 25(2), 139–145. <https://doi.org/10.1016/j.hrmr.2015.01.003>.
- Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, 15(3), e2000797. <https://doi.org/10.1371/journal.pbio.2000797>.
- Tan, J. Y., Huedo-Medina, T. B., Lennon, C. A., White, A. C., & Johnson, B. T. (2010). Us versus them in context: Meta-analysis as a tool for geotemporal trends in intergroup relations. *International Journal of Conflict and Violence*, 4(2), 288–297.
- Trzesniewski, K. H., & Donnellan, M. B. (2010). Rethinking “generation me”: A study of cohort effects from 1976–2006. *Perspectives on Psychological Science*, 5(1), 58–75. <https://doi.org/10.1177/1745691609356789>.
- Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Is “generation me” really more narcissistic than previous generations? *Journal of Personality*, 76(4), 903–918. <https://doi.org/10.1111/j.1467-6494.2008.00508.x>.
- Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, Connecticut: Graphics Press.
- Twenge, J. M. (1997a). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, 36(5–6), 305–325. <https://doi.org/10.1007/BF02766650>.
- Twenge, J. M. (1997b). Attitudes toward women, 1970–1995. *Psychology of Women Quarterly*, 21(1), 35–51. <https://doi.org/10.1111/j.1471-6402.1997.tb00099.x>.

- Twenge, J. M. (2000). The age of anxiety? Birth cohort change in anxiety and neuroticism, 1952–1993. *Journal of Personality and Social Psychology*, 79(6), 1007–1021. <https://doi.org/10.1037/0022-3514.79.6.1007>.
- Twenge, J. M. (2008). Generation me, the origins of birth cohort differences in personality traits, and cross-temporal meta-analysis. *Social and Personality Psychology Compass*, 2(3), 1440–1454. <https://doi.org/10.1111/j.1751-9004.2008.00094.x>.
- Twenge, J. M. (2009). Generational changes and their impact in the classroom: Teaching Generation Me. *Medical Education*, 43(5), 398–405. <https://doi.org/10.1111/j.1365-2923.2009.03310.x>.
- Twenge, J. M. (2017). *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood—and what that means for the rest of us*. New York: Simon and Schuster.
- Twenge, J. M., & Campbell, W. K. (2001). Age and birth cohort differences in self-esteem: A cross-temporal meta-analysis. *Personality and Social Psychology Review*, 5(4), 321–344. <https://doi.org/10.1207/S15327957PSPR05043>.
- Twenge, J. M., Campbell, S. M., Hoffman, B. J., & Lance, C. E. (2010). Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *Journal of Management*, 36, 1117–1142. <https://doi.org/10.1177/0149206309352246>.
- Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, 76(4), 875–902. <https://doi.org/10.1111/j.1467-6494.2008.00507.x>.
- Twenge, J. M., Zhang, L., & Im, C. (2004). It's beyond my control: A cross-temporal meta-analysis of increasing externality in Locus of Control, 1960–2002. *Personality and Social Psychology Review*, 8, 308–319. <https://doi.org/10.1207/s15327957pspr08035>.
- Tymula, A., Belmaker, L. A. R., Ruderman, L., Glimcher, P. W., & Levy, I. (2013). Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, 110(42), 17143–17148. <https://doi.org/10.1073/pnas.1309909110>.
- Urlick, M. J., Hollensbe, E. C., Masterson, S. S., & Lyons, S. T. (2017). Understanding and managing intergenerational conflict: An examination of influences and strategies. *Work, Aging and Retirement*, 3(2), 166–185. <https://doi.org/10.1093/workar/waw009>.
- Walters, N. T., & Horton, R. (2015). A diary study of the influence of Facebook use on narcissism among male college students. *Computers in Human Behavior*, 52, 326–330. <https://doi.org/10.1016/j.chb.2015.05.054>.
- Wegman, L. A., Hoffman, B. J., Carter, N. T., Twenge, J. M., & Guenole, N. (2018). Placing job characteristics in context: Cross-temporal meta-analysis of changes in job characteristics since 1975. *Journal of Management*, 44(1), 352–386. <https://doi.org/10.1177/0149206316654545>.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis (Vol. 59)*. Thousand Oaks, CA: Sage.
- Yang, Y. (2008). Age, period, cohort effects. In D. Carr (Ed.), *Encyclopedia of the life course and human development (Vol. 3, pp. 6–10)*. USA: Detroit: Macmillan Reference.
- Yang, Z., Cao, F., Lu, H., Zhu, X., & Miao, D. (2014). Changes of anxiety in Chinese military personnel over time: A cross-temporal meta-analysis. *International Journal of Mental Health Systems*, 8(19), 1–9. <https://doi.org/10.1186/1752-4458-8-19>.
- Yang, Y., & Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, 36(1), 75–97. <https://doi.org/10.1111/j.1467-9531.2006.00175.x>.
- Yang, Y., & Land, K. C. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications*. Boca Raton, FL: CRC Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.