Measurement Discrepancies Between Logged and Self-Reported Digital Media Use: A Systematic Review and Meta-Analysis

Douglas A. Parry,[a] Brittany I. Davidson,[b, c] Craig J. R. Sewall,[d] Jacob T. Fisher,[e] Hannah Mieczkowski,[f] Daniel S. Quintana[g, h]

[a]Department of Information Science, Stellenbosch University, South Africa

[b]School of Management, University of Bath, UK

[c]Department of Computer Science, University of Bristol, UK

[d]School of Social Work, University of Pittsburgh, USA

[e]College of Media, University of Illinois Urbana-Champaign, USA

[f]Department of Communication, Stanford University, USA

[g]NORMENT, Division of Mental Health and Addiction, University of Oslo, and Oslo University Hospital, Oslo, Norway

[h]KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway

Author Notes:
This manuscript is a pre-print (the authors' original version) that has not yet received peer review. Please cite responsibly.

This manuscript is structured as follows: Abstract ⋄ Introduction ⋄ Results ⋄ Discussion ⋄ Methods ⋄ References ⋄ Supplementary Material

Funding Declaration: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests: The author(s) declare that there are no conflicts of interest with respect to the authorship or the publication of this article.

Acknowledgements: Thank you to all of the researchers who made their data available for inclusion in this review. We are especially grateful for those who completed extra analyses to provide us with relevant data and statistics. Additionally, we acknowledge the role played by those who shared our public calls for relevant data.

Correspondence concerning this article should be addressed to Douglas Parry, Department of Information Science, Stellenbosch University, Stellenbosch, South Africa; Email: dougaparry@sun.ac.za

Abstract

The influence of digital media on personal and social well-being is a question of immense public and academic interest. Scholars in this domain often use retrospective self-report measures of the quantity or duration of media use as a proxy for more objective measures, but the validity of these self-report measures remains unclear. Recent advancements in log-based data collection techniques have produced a growing collection of studies indexing both self-reported media use and device-logged measurements. Herein, we report a meta-analysis of this body of research. Based on 106 effect sizes, we found that self-reported media use was only moderately correlated with device-logged measurements, and that these self-report measures were rarely an accurate reflection of logged media use. These results demonstrate that self-reported measures of the quantity or duration of media use are not a valid index of the amount of time people actually spend using media. These findings have serious implications for the study of media use and well-being, suggesting that cautiousness is warranted in drawing conclusions regarding media effects from studies relying solely on self-reported measures of media use.

Measurement Discrepancies Between Logged and Self-Reported Digital Media Use: A Systematic Review and Meta-Analysis

The widespread adoption of various digital media (e.g., smartphones, games, social media, etc.) has brought an immense public and academic interest in understanding how media use impacts an individual's psychological and social well-being (Dickson et al., 2019). Some conclude that media use has "*destroyed a generation*" (e.g., Twenge, 2017). Others decry these claims, noting that media use is no more strongly associated with harmful outcomes than is potato consumption or wearing glasses (Orben & Przybylski, 2019), and suggesting that current concern regarding the potentially negative influence of media on well-being is merely this generation's manifestation of a "*Sisyphean Cycle of Technology Panics*" (Orben, 2020). This incongruity is not only problematic for knowledge generation but, because findings have the potential to foment far-reaching societal or policy changes (e.g., a bill has been introduced in the US Senate aimed at limiting the time users can spend on digital media platforms to 30 minutes a day; S. 2314, 2019), it may actually be harmful for individuals and society.

Progress towards resolving this debate and developing a deeper understanding of the role of media use in human behaviour and well-being requires "*transparent and robust analytical practices*" (Orben & Przybylski, 2019), but also confidence that the measurement tools used to assess media use are valid indicators of actual usage patterns (John & Benet-Martínez, 2014; Flake et al., 2017). Before conclusions can be made about the effects of media use, we must first trust not only the theoretical models posed in studies, but also the measures used to produce data to test these models. At present the majority of research in this domain relies on retrospective self-report measures to quantify media use, and to infer the relationship between usage patterns and individual or group-level differences in well-being (Gerpott & Thomas, 2014; Howard & Jayne, 2015; Griffioen et al., 2020).

Self-report measures for media use typically concern either: 1) the time spent using all media, 2) time spent using specific media, or 3) the frequency or volume of media use (Jungselius & Weilenmann, 2018). Responses are usually collected in the form of single point estimates or Likert-type scales. Accurate estimation is affected, not only by well-known factors that affect survey-response behaviour (e.g., instructions, response option formatting, question-order, response strategies, reference periods, memory, motivation, desirability biases, or affective states; Jobe, 2003; Tourangeau et al., 1984), but also by the fact that media use is likely to be especially difficult to report accurately. People often use multiple media simultaneously and embed media use alongside other non-media activities (e.g., texting while eating or browsing the web in a lecture; Yeykelis et al., 2014; Brasel & Gips, 2017). In addition, media use frequently consists of numerous micro-interactions (Andrews et al, 2015), further blurring the distinction between media and non-media activities (Vorderer et al., 2016). Taken together, given known difficulties estimating behaviours that are highly integrated into respondents' lives (Jobe, 2003), media use is likely to be difficult to recall and to accurately estimate, undermining the validity of self-report measures and the produced on their basis.

In addition to using self-report measures of media use quantity or duration, researchers frequently use self-report measures of *problematic* media use (e.g., technology-related addictions or other conceptions of problematic media use) as proxies for general use (Davidson et al., 2020b; Ellis et al., 2019; Ellis, 2019). These problematic media use scales frequently include items that concern cognitive or affective responses to, or motivations for media use alongside items assessing media use quantity or duration (Abendroth et al., 2020). As such, recent work suggests that these measures may be more representative of general traits than of media use (Ellis, 2019, p. 61).

Over the preceding decade, adoption of data-intensive approaches for measuring media use has accelerated. In parallel with general developments in personal analytics have come tools that enable researchers to directly measure usage behaviour across a variety of media (Piwek et al., 2016; Ryding & Kuss, 2020). Along with network or operator level traffic logging techniques, these research methods also harness operating system capabilities or proprietary and third-party applications built for consumer audiences. Such methods enable researchers to measure complete device use, network or call traffic, or even the use of specific applications and services (Gerpott & Thomas, 2014).

These developments have led to a number of investigations considering associations between self-reported and logged media use. Early research in this domain showed that for calling and texting on mobile phones, self-reports correlate only moderately with network provider logs (Boase and Ling, 2013; Vanden Abeele, 2013). Comparisons between digital trace data of Internet use and self-reported use indicate similarly moderate correlations (Scharkow, 2016). Recently, Ellis et al. (2019) compared responses for ten scales and three single estimates for either general or problematic use of smartphones with relevant tracking data. While all self-report measures positively correlated with device use, effect sizes were small—a pattern that seems to hold across a number of studies (Andrews et al., 2015; Boase & Ling, 2013; Ellis et al., 2019; Scharkow, 2016; Vanden Abeele et al., 2013).

These data suggest that self-reported and device-logged media use measures, rather than serving as different ways to measure media use, may in fact capture distinct constructs (Kobayashi & Boase, 2012; Sewall et al., 2020). As such, there exists a need to systematically assess whether self-reported media use is an accurate indicator of actual usage patterns. Evidence of a weak association between these two measures would suggest that the validity of the current body of evidence for media effects, and the associated policy recommendations, is questionable. To address this gap in knowledge, we conducted a pre-registered systematic review and meta-analysis of extant research wherein both self-reported and device-logged media use were assessed. In addition, we assessed the extent to which the relationship between self-reported and logged use depends on the medium or the characteristics of the measure, and whether individuals tend to systematically under- or over-report their media use.

Results

Included effect sizes. The systematic search produced 12,132 results. After screening for eligibility (see Figure 1), 47 records were included in the final sample. From these records, 106 effect sizes were included in the analyses, with a combined sample size of n = 57,559. Supplementary Table 1 provides a summary of the included effect sizes for measures concerning use and Supplementary Table 2 provides a summary for measures concerning problematic use.

To evaluate the association between self-reported and logged media use, 66 effect sizes from 44 studies were considered. Across these comparisons the total sample size was N = 52,007. On average, a comparison involved 787.99 participants (SD = 1621.27, median = 166, min = 20, max = 6,598). In a second meta-analysis we investigated associations between self-reported problematic media use and logged measures of media use. This analysis included 42 effect sizes from 19 studies, with a total sample size of N = 5,552. On average, a comparison involved 138.8 participants (SD = 92.79, median = 139.5, min = 14, max = 294). Finally, to assess whether individuals tend to systematically under- or over-report their media use, we included 49 comparisons from 30 studies and a total sample size of N = 17,523, with an average sample size of 357.61 participants (SD = 955.62, median = 159, min = 20, max = 6,598).

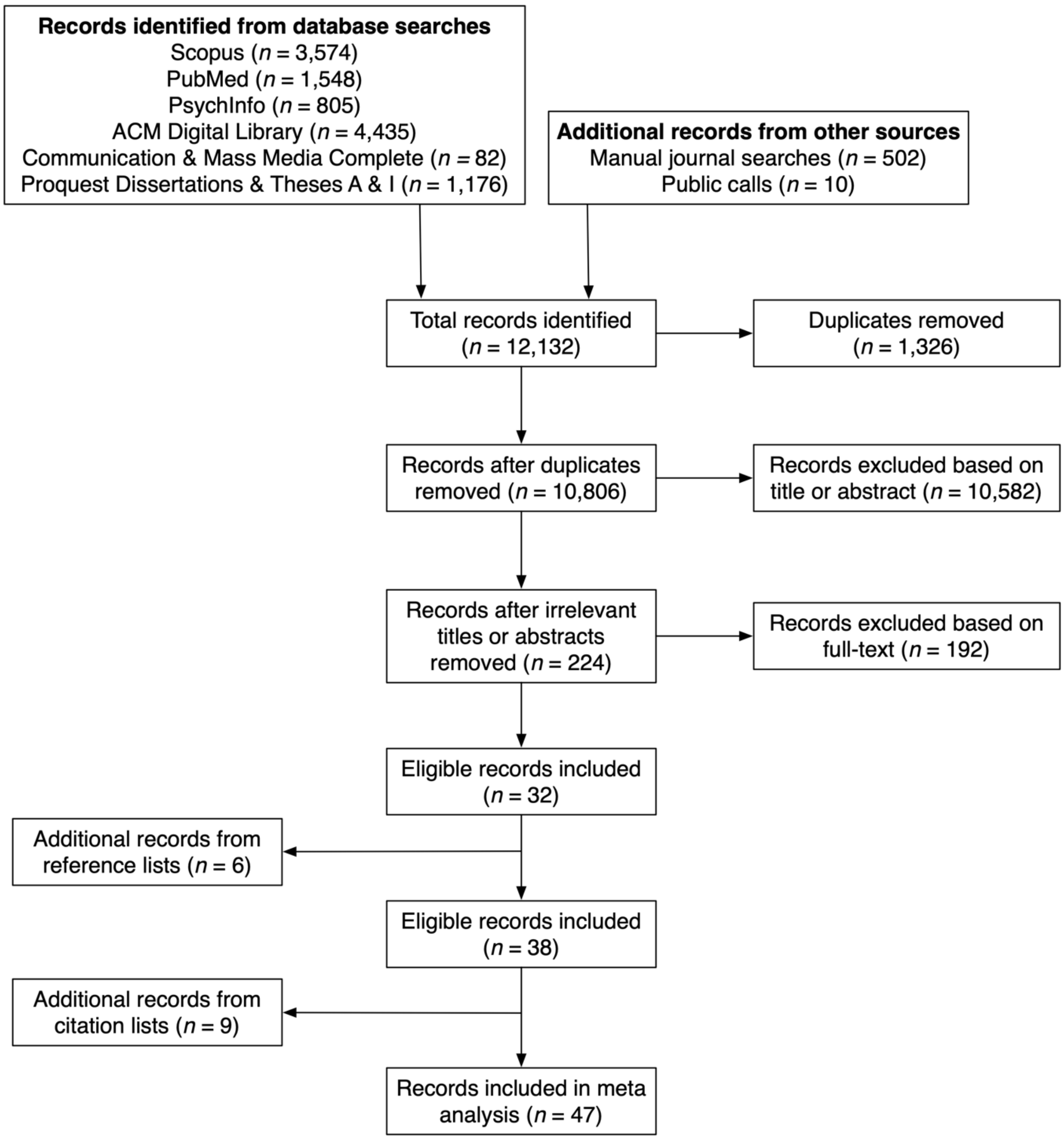

Figure 1. Flow diagram for the inclusion process.

Correlations between self-report and logged measures

Associations for media use self-reports. The correlation between self-reported and logged measures of media use was calculated with robust variance estimation — RVE, revealing a relationship that was positive, but only medium in magnitude ($r = 0.38$, 95% CI [0.33, 0.42], $p < 0.001$). Figure 3 depicts a forest plot of the effect sizes included in this analysis. Egger's regression test (incorporating RVE per Rodgers and Pustejovsky (2020)'s Egger Sandwich test), indicated no evidence of small study bias in this sample ($\beta = 0.53$, $p = 0.143$; see Panel A in Figure 2 for a contour-enhanced funnel plot).

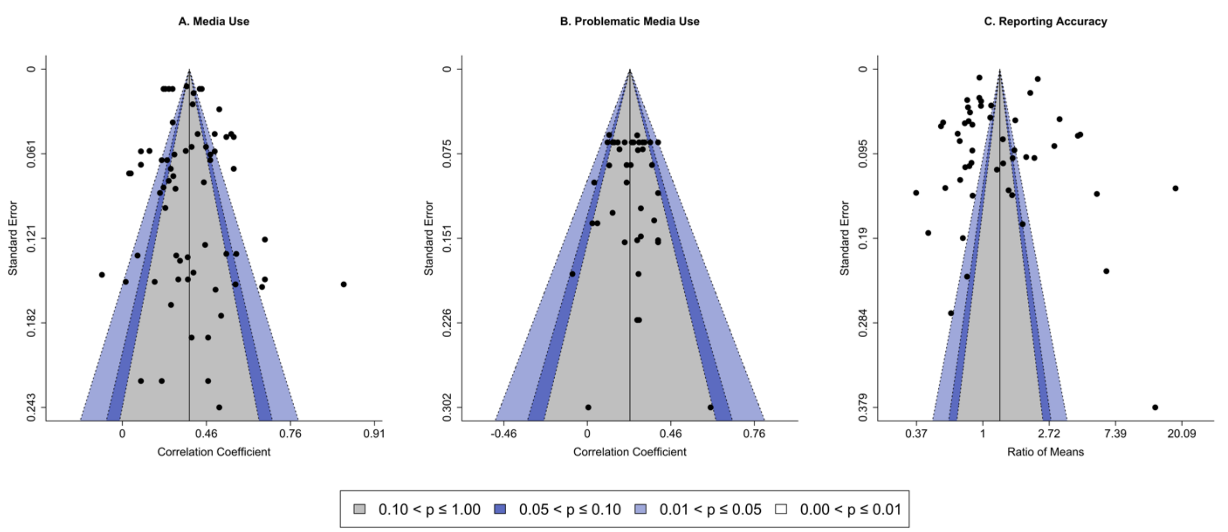

Figure 2. Contour-enhanced funnel plots depicting the relationship between the observed effect sizes (on the x-axis) and their standard errors (on the y-axis) for comparisons concerning media use (A), problematic media use (B) and reporting accuracy (C). The vertical lines indicate the estimated summary effect size. The shaded bands represent the significance contours indicated in the legend and each black dot represents an observed effect size. Visual inspection of all three plots does not indicate asymmetry, nor does it indicate evidence of publication bias as there is no obvious overrepresentation of effect sizes in the highlighted significance contours.

Influence diagnostics indicated a single outlier in this sample (Junco, 2013; $n = 45$, $r = 0.87$). A sensitivity analysis excluding this outlier produced a summary effect size that was almost the same as the original analysis ($r = 0.37$, 95% CI [0.33, 0.42], $p < 0.001$). Similarly, a sensitivity analysis excluding the only effect size in the analysis that was extracted using a web plot digitiser (Katapally and Chu, 2019, $n = 49$, $r = -0.12$) showed a comparable effect size to the original analysis ($r = 0.38$, 95% CI [0.34, 0.43], $p < 0.001$). As these two analyses indicated only minor effects of the comparisons in question on the meta-analysis, they were retained. As a final sensitivity analysis, we considered whether the results presented in peer-reviewed studies differed from non-peer reviewed studies. Of the 66 included effect sizes, 10 (15.15%) were non-peer-reviewed at the time of inclusion (see Supplementary Table 1). While the effect size is larger in peer-reviewed ($r = 0.39$, 95% CI [0.34, 0.44], $p < 0.001$, $k = 56$) than in non-peer-reviewed ($r = 0.32$, 95% CI [0.21, 0.42], $p = 0.001$, $k = 10$) effects, the difference is not statistically significant ($\beta = -0.08$, 95% CI [-0.21, 0.04], $p = 0.169$).

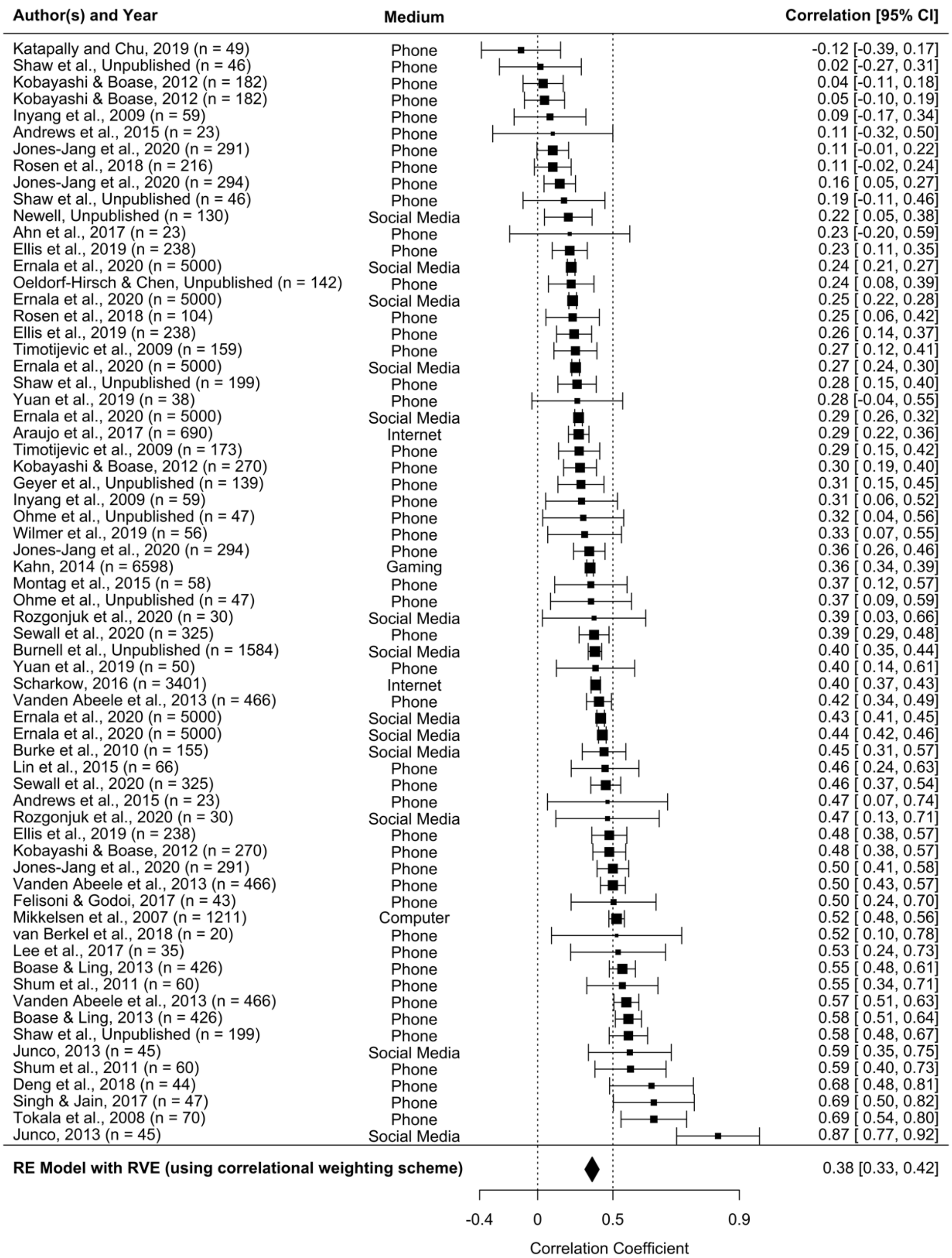

Figure 3. Forest plot of the effect sizes for studies included in the meta-analysis for the association between self-reported and logged measures of digital media use. The dashed reference line at the intercept for $r = 0.5$ represents the point from which the magnitude of the association would be sufficient to conclude that the measures are appropriate substitutes for one another (Carlson & Herdman, 2012). Error bars represent 95% confidence intervals of the effect size. RE = Random effects model. RVE = Robust variance estimation (conducted with a correlated effects weighting scheme).

Associations for problematic media use self-reports. The correlation between self-reported problematic media use and logged media use (calculated with RVE) was positive, but small ($r = 0.25$, 95% CI [0.20, 0.29], $p < 0.001$), with a low level of heterogeneity ($Q(41) = 59.85$, $p = 0.017$; with RVE: $T^2 = 0.003$, $I^2 = 28.51\%$). Figure 4 presents a forest plot for this analysis. Egger's regression test (incorporating RVE per Rodgers and Pustejovsky, 2020), indicated no evidence of small study bias ($\beta = 0.34$, $p = 0.244$; see Panel B in Figure 2 for a contour-enhanced funnel plot). Because influence diagnostics did not reveal any outliers, a leave-one-out analysis was not conducted. Similarly, as no included effects were extracted using the web plot digitiser, no sensitivity analysis was conducted. However, because five included effects were presented in non-peer-reviewed studies, we considered whether this influenced the outcome. For peer-reviewed studies the correlation was estimated with RVE while, for non-peer-reviewed studies, there were insufficient observations so a random-effects intercept-only model was calculated. No meaningful difference was observed between peer-reviewed ($r = 0.25$, 95% CI [0.19, 0.30], $p < 0.001$, $k = 35$) and non-peer-reviewed ($r = 0.25$, 95% CI [0.15, 0.34], $p < 0.001$, $k = 5$) effects ($Q_b(1) = 0.01$, $p = 0.934$).

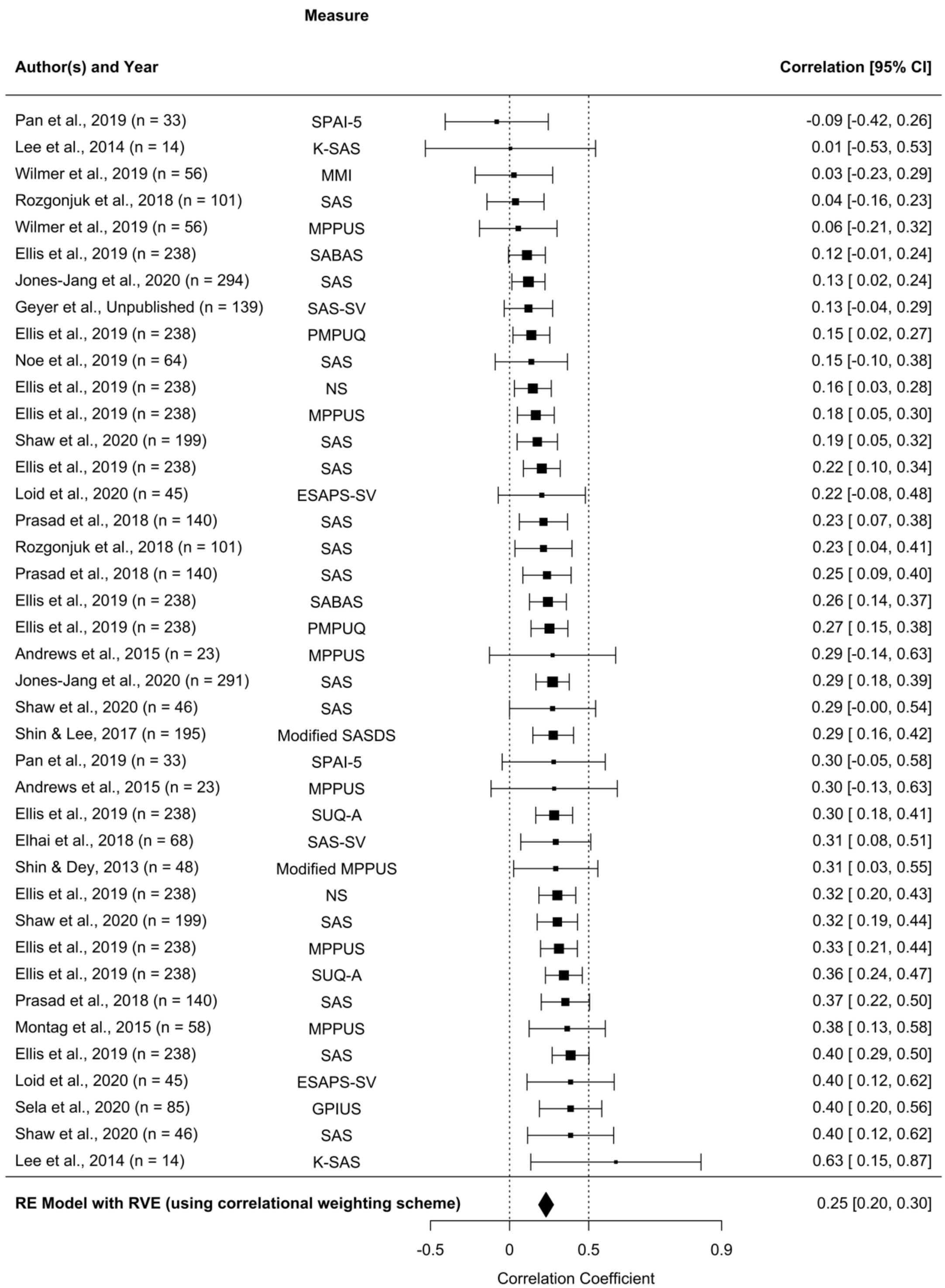

Figure 4. Forest plot of the effect sizes for studies included in the meta-analysis for the association between self-reported problematic media use and logged measures of media use. The dashed reference line at the intercept for *r* = 0.5 represents the point from which the magnitude of the association would be sufficient to conclude that the measures are appropriate substitutes for one another (Carlson & Herdman, 2012). Error bars represent 95% confidence intervals of the effect size. RE = Random effects model. RVE = Robust variance estimation (conducted with a correlated effects weighting scheme).

Moderators of the association between self-reported and logged media use

There was a high level of heterogeneity in the included effect sizes ($Q(63) = 752.22$, $p < 0.001$; with RVE: $T^2 = 0.012$, $I^2 = 92.42\%$) for the correlation between self-reported and logged media use. Therefore, as planned, three moderator analyses were conducted to try to identify possible sources of heterogeneity (moderator analyses were not specified for comparisons involving problematic usage). While sufficient data were available for self-report form (Scale: $k = 6$; Estimate: $k = 60$) and self-report category (Duration: $k = 47$; Volume: $k = 19$), only two levels for medium (Phone: $k = 49$; Social media: $k = 13$) met our requirements. Therefore, deviating from our analysis plan, we only considered summary effect sizes for studies investigating use of phones or social media.

Type of Media. We first calculated a summary effect size for studies targeting use of a phone or social media and found it to be comparable to the overall correlation ($r = 0.38$, 95% CI [0.32, 0.42], $p < 0.001$). While the association was smaller for social media ($r = 0.36$, 95% CI [0.27, 0.44], $p = 0.001$) than for phones ($r = 0.39$, 95% CI [0.31, 0.45], $p < 0.001$), this difference was not statistically significant ($\beta = -0.02$, 95 % CI [-0.14, 0.09]. $p = 0.672$).

Self-report Form. While the small number of studies using scales ($k = 6$) impacts interpretability, we found that the difference in the magnitude of the association between scales ($r = 0.25$, 95% CI [-0.004, 0.47], $p = 0.052$) and single estimates ($r = 0.39$, 95% CI [0.34, 0.43], $p < 0.001$) was not statistically significant ($\beta = 0.13$, 95% CI [-0.18, 0.42], $p = 0.300$).

Self-report Category. We found no evidence that the association between self-reported and logged measures of media use differs between measures concerning either the duration (e.g., usage time; $r = 0.38$, 95% CI [0.33, 0.43], $p < 0.001$) or the volume (e.g., usage amount; $r = 0.34$, 95% CI [0.25, 0.43], $p < 0.001$) of use ($\beta = -0.001$, 95% CI [-0.13, 0.13], $p = 0.979$).

Accuracy of self-report measures

Analysis of aggregated accuracy. Of the 49 included comparisons, only three (6.12%) mean self-reported media use estimates fell within 5% of the logged mean. Despite this, similar proportions of studies reported mean self-reports of media use that were either over- (k = 23, 46.94%) or under- (k = 23, 46.94%) reported relative to the logged measure.

To produce a summary effect size, we calculated the weighted ratio of means (incorporating RVE after log transformation) between self-reported and logged measures of media use and found that, across studies, participants over-reported their media use (R = 1.21, 95% CI [0.94, 1.54], p = 0.129). However, given that the confidence interval for this result includes indicator values for under-reported and accurately reported media use, the evidence is insufficient to conclude whether estimates are typically under- or over-reported compared to logs of media use. Figure 5 provides a forest plot for the effects included in this analysis.

For this sample, Egger's regression test (incorporating RVE per Rodgers and Pustejovsky, 2020) showed no evidence of small study bias (β = 0.61, p = 0.41; see Panel C in Figure 2 for a contour-enhanced funnel plot). Influence diagnostics indicated a single outlier (Junco, 2013; n = 45, r = 0.87, self-report mean = 73 minutes, self-report SD = 59, logged mean = 4 minutes, SD = 6; R = 18.25, 5% CI [14.05, 23.71]). A sensitivity analysis excluding this outlier produced a summary effect size that was similar to the original analysis (R = 1.18, 95% CI [0.95, 1.48], p = 0.136). Of the 49 effects, nine (18.37%) were non-peer-reviewed at the time of inclusion (see Supplementary Table 1). A sensitivity analysis excluding these studies found no statistically significant difference between peer-reviewed (R = 1.30, 95% CI [0.97, 1.75], p = 0.075) and non-peer-reviewed (R = 0.89, 95% CI [0.57, 1.40], p = 0.543) effects (β = -0.367, *Exp*(β) = 0.69, 95% CI [0.41, 1.16], $p = 0.133$). A second sensitivity analysis excluding two effects that were included after using the web plot digitiser (Inyang et al., 2009; Lee et al., 2017) showed comparable results to the overall analysis (R = 1.21, 95% CI [0.94, 1.56], $p = 0.141$).

Moderators of reporting accuracy. There was a high-level of heterogeneity in the sample ($Q(48) = 7222.29$, $p < 0.001$; with RVE: $T^2 = 0.32$, $I^2 = 99.49\%$). Two moderator analyses were planned a priori to investigate possible sources of heterogeneity. For medium, only two levels (Phone: $k = 41$; Social Media: $k = 5$) held sufficient data. For the self-report category, there was sufficient data for measures of duration ($k = 35$) and volume ($k = 14$).

Type of Media. For studies including both self-report and logged measures of phone use, the summary effect size for this subgroup was comparable to the overall analysis (R = 1.07, 95% CI [0.85, 1.35], $p = 0.574$). For social media (R = 2.89, 95% CI [0.18, 46.04], $p = 0.241$), the Satterthwaite degrees of freedom for the model were less than 4, indicating a high probability of a Type I error and that the results should not be trusted (likely due to the small number of effects and the presence of an extreme outlier). Consequently, no moderator analysis was conducted.

Self-report Category. While measures of duration showed a larger degree of over-reporting (R = 1.29, 95% CI [1.01, 1.66], $p = 0.044$) compared to measures of volume which indicated under-reporting (R = 0.80, 95% CI [0.57, 1.11], $p = 0.162$), the difference was not statistically significant ($\beta = -0.44$, *Exp*(β), = 0.64, 95% CI [0.41, 1.02], $p = 0.056$).

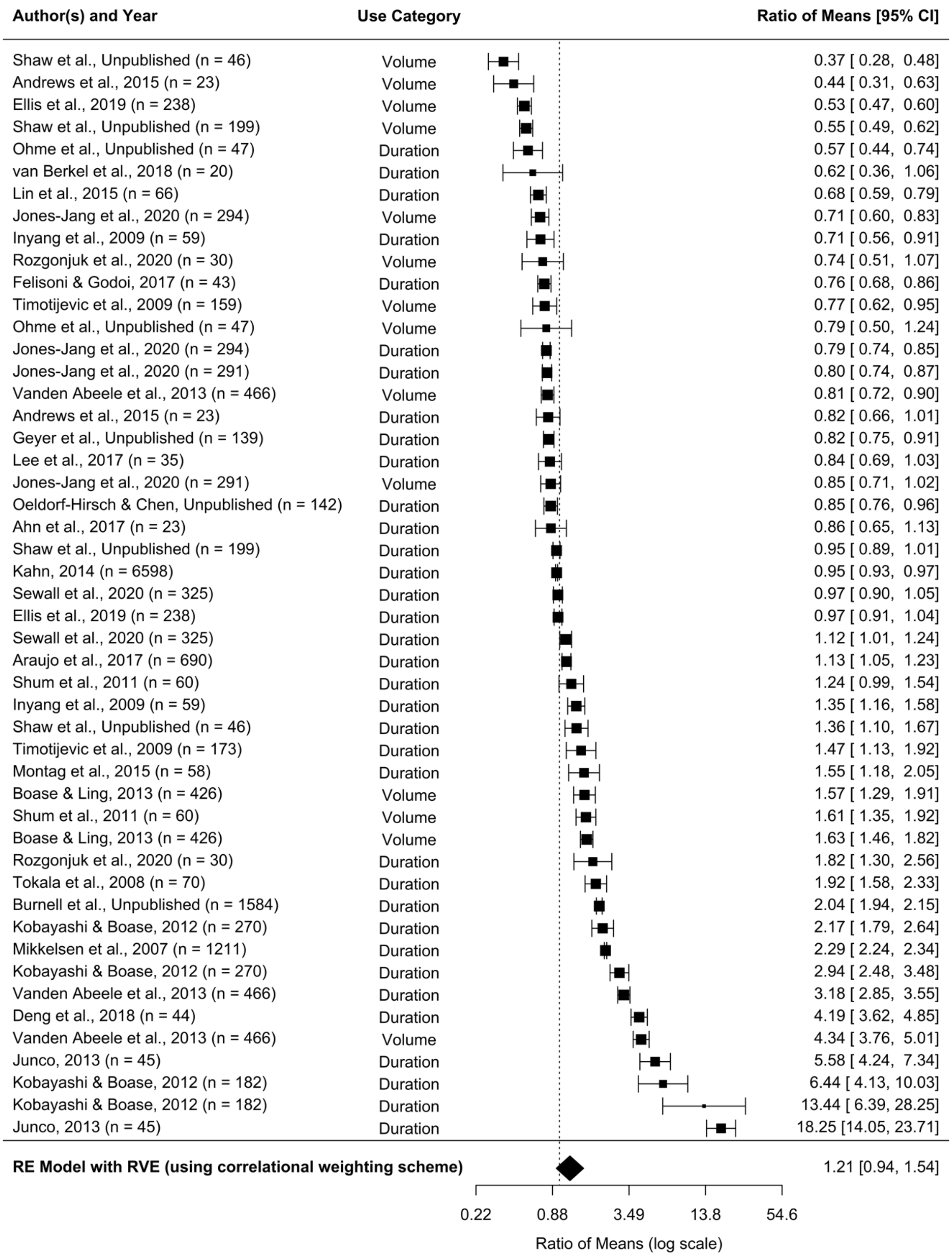

Figure 5. Forest plot of the effect sizes for studies included in the meta-analysis for the ratio of means between self-reported and logged measures of digital media use represented on a log scale. The reference line at the intercept for 1.0 represents a 1:1 ratio between self-reported and logged media use, with values below one indicating under-reporting and values above one indicating over-reporting of media use.

Discussion

Over the preceding decade, both academic and popular press articles have suggested that digital media use may be responsible for increases in depression, anxiety, loneliness, and suicidality, among other negative effects on well-being—especially among adolescents (e.g., Chuck, 2017; Twenge, 2017). In response to this research, wide-ranging policy recommendations, interventions, and legislation have been instituted in efforts to protect youth from the perceived harms of media use (e.g., Malinauskas & Malinauskiene, 2010; Newton, 2019; American Academy of Pediatrics, 2016). The findings of these studies, however, are largely gleaned from self-report data. Our findings indicate only a modest association between these self-report measures and device logs of actual media use ($r = .38$). For problematic use scales, this association was even smaller ($r = .25$).

While there is no widely accepted threshold for convergent validity (Nunnally, 1994; Carlson & Herdman, 2012), given the magnitude of the associations found in this meta-analysis, the available evidence suggests that self-reported measures of media use should not be considered suitable substitutes for logs of media use. Furthermore, our observation of an even smaller association between problematic use scales and device logs suggests even more caution when adopting measures of problematic media use as proxies for usage itself (Ellis, 2019; Davidson et al., 2020a). Moreover, while the results show that similar proportions of studies indicate either under- or over-reporting, less than 10% of self-reports were within 5% of the equivalent logged value, indicating that, when asked to estimate their usage, participants are rarely accurate.

If these measures are not valid representations of the behaviour they are intended to assess, there is uncertainty about the nature of what is being measured (Davidson et al., 2020b).

Perhaps more concerningly, this uncertainty extends to the findings and conclusions provided by studies using these measures to quantify media use (Cronbach & Meehl, 1955; John & Benet-Martínez, 2014; Flake et al., 2017; Flake & Fried, 2020), raising immediate implications for media effects research. Given substantial academic and public interest in the putative effects of media use, research in this area can have a major impact on popular opinion, policy recommendations, and interventions. For example, the article "*Have Smartphones Destroyed a Generation?*" (Twenge, 2017) has been shared over 850,000 times on social media (as indicated by *BuzzSumo* on 25 September 2020). The results presented herein suggest pause in drawing wide-reaching conclusions related to policies and interventions from purely self-report studies.

Given the predominance of self-report measures and the use of correlational designs and statistical analyses (Scharkow, 2016; Ellis, 2019), the effect of measurement error requires consideration. Some studies provide support for the argument that self-reports have attenuated effect sizes and increased the likelihood of false negatives (Jones-Jang et al., 2020), a larger number of studies, however, suggest that the (in)accuracy of self-reported media use measures may be systematic. For instance, multiple studies have found that the accuracy of self-reported media use depends, in part, on how much the respondent uses media (Araujo et al., 2017; Scharkow, 2016; Sewall et al., 2020; Vanden Abeele et al., 2013). Furthermore, a recent study (Sewall et al., 2020) found that the degree of inaccuracy was directly related to the respondent's level of well-being. In a similar manner, evidence suggests that there is substantial conceptual overlap between various mental health constructs (e.g., anxiety, stress, depression) and problematic media use scales, but little (if any) of the variance explained includes usage (Davidson et al., 2020b). Taken together, this suggests that self-reported estimates of media use are likely systematically biased by the very factors that are fundamental to the associations under investigation. While our meta-analysis has shown that, across studies, the association between logged and reported media use is generally insufficient to conclude that the measures are appropriate substitutes, given the information reported in primary studies, further investigation is needed to investigate the likely systematic nature of this discrepancy.

Although the correlations observed are indicative of poor convergent validity, there remains a high-level of heterogeneity in effect sizes for correlations involving self-reported usage. Similarly, the ratio of means between logged and self-reported media use displayed a high level of heterogeneity, with reports ranging from under-reporting to over-reporting. Taken together, these outcomes indicate that the observed association and degree of over-reporting may not be consistent. Various methodological, contextual, participant, or medium-specific factors may impact the magnitude of associations between self-reports and logged measures of media use. To investigate this heterogeneity, we considered whether findings were influenced by relevant measurement factors. The results, however, indicate that the correlation is not moderated by self-report form, nor is it moderated by the category of use. Similarly, self-report accuracy was not moderated by whether estimates concerned the volume or duration of media use. For both usage correlations and response-accuracy, our investigation of the moderating effect of different media was hampered by the absence of a sufficient number of studies measuring both logged and self-reported use within each category. For this reason, these results cannot definitively speak to the moderating effect of medium on the relationship between self-reported and logged measures of media use. Notwithstanding these assessments, the remaining unexplained heterogeneity in associations between logged and reported media use, and the degree to which participants accurately estimate their usage, are important avenues for future research. In contrast to these two assessments, only a low level of heterogeneity was observed for correlations involving self-reported problematic media use. This suggests, firstly that the relationship is relatively stable across comparisons and, secondly, given the differences in observed correlations and heterogeneity between general usage self-reports and problematic usage self-reports, that measures of problematic media use capture constructs distinct from those reflected in general media use self-reports (Davidson et al., 2020b).

Limitations and implications for future research

This study is not without limitations. First, although a number of analyses were conducted to assess potential biases, there remains the possibility that various publication biases may have had an impact on the targeted literature base potentially influencing our study outcomes. Second, while moderator analyses were conducted to investigate possible sources of heterogeneity, many moderator levels included too few effects for meaningful analysis and interpretation. Specifically, as too few studies have considered associations between self-reported and logged use of specific media such as computers, games, or social media, we are unable to conclusively evaluate the degree to which the relationship between recorded and estimated media use differs across various media. Similarly, while included assessments of problematic media use only used scales, self-reported media use mostly occurred through single estimates. Consequently, to extend the moderator analysis conducted, further evidence is needed to determine whether the form of a self-report measure affects the alignment between self-reported and logged measures of media use. Acknowledging the remaining heterogeneity, and the possibility that various participant-level characteristics may impact associations (Kobayashi & Boase, 2012; Sewall et al., 2020), there is a need to further investigate the factors that might systematically account for the discrepancy observed between self-reported and logged measures of media use. While well-known factors affecting survey-response behaviour may account, to some extent, for inaccuracies, there is a need to investigate whether specific user (e.g.., behavioural or personal characteristics) or technological (e.g., medium or device) characteristics are involved.

An additional factor that may explain some of the heterogeneity in the observed associations lies in the usage logging methodology adopted in the included studies. As is evident in Supplementary Table 1 and Supplementary Table 2, a variety of proprietary, custom, and consumer-level tools have been used to log media use. Notwithstanding the need for methodological progress, although poor convergent validity is indicative of weak construct validity, it is not sufficient— "*when convergent validity is weak, one or both variables do not capture the intended construct well*" (Carlson & Herdman, 2012, p. 20). While at face-value tracking methods provide accurate and valid measurements of media use, the possibility of biases and inaccuracies cannot be ignored (Scharkow, 2016). Irrespective of validity concerns, automatically logged data is not a panacea for media use measurement (Jürgens et al., 2019; Ryding & Kuss, 2020). A particular challenge for widespread adoption of such methods relates to sampling biases occurring due to differential willingness and abilities to provide such data. Moreover, in addition to technical incompatibilities (device or system restrictions, gaps in coverage), participant biases (reactivity), and increased resource demands (time, cost, and participant burden), there are substantial ethical, security and privacy related challenges associated with tracking media use (Jürgens et al., 2019; Scharkow, 2016). Such biases and challenges, while not necessarily affecting the accuracy of the data produced, nonetheless, merit careful consideration. Furthermore, just as calls for higher standards of evidence have prompted examination of the validity of self-report measures of media use, there is a need to regularly reflect on the validity of logged measures and, in parallel, continually develop improved tools for quantifying media use (Sen et al., 2019).

Our findings contribute to ongoing efforts to move research beyond simple conceptions of ‘screen-time’ (Kaye et al., 2020) and develop more nuanced conceptions of the role of media in human behaviour. As Orben (2020, p. 6) notes, while there do exist applicable theories in this domain, with few exceptions, they are used in a superficial manner and are not a driving force progressively shaping a cumulative research tradition. The conceptual tension brought about by these validity concerns can stimulate a drive for theories that have a higher degree of verisimilitude and greater utility for addressing important questions facing society today. In addition to the need for research on digital media use and effects to systematically move on from “the repetitive development of self-report assessments” (Ellis et al., 2019, p. 13), beyond the focus of this investigation, as Kaye et al. (2020), Meier and Reinecke (2020), and Büchi (2020) discuss, there is a profound need for a paradigm shift in which the specific affordances enacted, and the behaviours, digital practices, experiences, and outcomes that they facilitate, receive central focus, rather than simply the overall duration or volume of usage. Coupled with more valid measures and transparent and robust analytical practices, such developments will bring us closer to understanding the role of digital media in human behaviour and psychological well-being.

Method

As described in Quintana (2015) we developed a methodology for the systematic search and meta-analysis. To pre-register our expectations and methodology, our systematic review protocol, developed in-line with the PRISMA-P guidelines (Moher et al., 2015), was made publicly accessible (Parry et al., 2020) prior to data collection. All materials required to reproduce the results of the study are available on the Open Science Framework (OSF: https://osf.io/dhx48/).

To guide our review, we pre-specified a number of relevant predictions and questions. These are presented as formal hypotheses and research questions in our study protocol. However, for the sake of brevity, here we simply provide an overview of our a priori predictions for the meta-analysis, before outlining the details of our data collection and analysis procedures.

Given the accuracy and validity issues with self-report measures of media use, we expected the association between self-reported measures of media use and measures produced from digital trace data to be positive, but only small to medium in magnitude. To separate the comparison of measures of general media use from measures of problematic media use, to address this first expectation, we only considered comparisons between self-report measures that concern either the total or average *duration* (e.g., minutes, hours) or *volume* (e.g., number of pickups, number of logins, number of phone calls etc.) of media use and equivalent logged measures for the same period (e.g., daily, weekly etc.). To understand if the association between self-reports and logged measures is affected by characteristics of the medium or the self-report measure we explored whether it is moderated by (a) the medium (i.e., social media, smartphones, the Internet, computers, gaming), (b) the form of self-report measure (i.e., a single estimate or a scale), or (c) the category of media use (i.e., volume of interactions or duration of usage).

In addition to considering associations between measures explicitly concerning media usage, acknowledging that, despite concerns over validation procedures (Harris et al., 2020; Laconi et al., 2014) and questionable relations between the constructs assessed and use (Davidson et al., 2020b), scales assessing problematic media use (i.e., technology-related addictions or other conceptions of problematic use) are frequently adopted to assess media use

(Ellis, 2019; Shaw et al., 2020), we investigated the association between such measures and logged measures of media use. For this separate analysis we also expected the association between self-reported measures of problematic media use and measures produced from digital trace data to be positive but small to medium in magnitude.

Our final aim concerned the accuracy of self-report measures, relative to equivalent logged measures of media use. To this end, we assessed whether participants typically under- or over-report their media use compared to equivalent logged measures. To understand if there are factors that systematically affect accuracy, we investigated if there is evidence indicating that measurement error is systematically related to either the medium or the category of media use involved in a comparison.

Search Strategy and Selection Criteria

To identify relevant published studies, we conducted an automated search on five broad bibliographic databases: *PubMed*, *Scopus*, *PsychInfo*, *Communication & Mass Media Complete,* and the *ACM Digital Library*. To target unpublished (grey) literature we used the *ProQuest Dissertations & Theses A&I* database. A generic search string was developed in consultation with an academic librarian at Stellenbosch University and, for each database, was adjusted as required. The search string includes four clauses, with at least one matching term required for each clause. The first clause includes terms relating to various forms of eligible media (e.g., social media OR Internet OR phone OR games, etc.). The second and third clauses relate to logged data (e.g., server logs OR track, etc.) and self-report measures (e.g., survey OR self-report OR questionnaire, etc.), respectively. The fourth clause includes terms relating to media use (e.g., use OR usage OR behaviour, etc.). The full search strings (applied to the title, abstract, and keywords fields or just the abstract field if restricted) are available through the OSF (https://osf.io/dhx48/). In addition to the automated search, a manual search was conducted within five relevant journals (*Human Communication Research; Cyberpsychology, Behavior and Social Networking; Communication Methods and Measures; International Journal of Human-Computer Studies; Media Psychology*). Following assessment for eligibility, the included studies were supplemented by ‘backward’ and ‘forward’ search procedures (Webster & Watson, 2002) using the *Google Scholar* search engine. Finally, we made public calls for relevant unpublished data and papers.[1]

We restricted inclusion to studies that collected both self-reported and logged measures of media use. For self-reports, eligible scales or single estimates should have either concerned use in general (i.e., volume or duration) or problematic use (i.e., technology-related addictions or other conceptions of problematic use). These self-report and logged measures for media use should have concerned use of either social media, games, a mobile phone, the Internet in general, or a computer. In addition to these criteria, we restricted inclusion to studies published since 2007 (inclusive), the initial release year for the iOS operating system (with the release of Android in the following year), and a time from which use of social networking services gained widespread popularity. We restricted inclusion to studies reported in English. While we excluded studies that explicitly targeted clinical populations, no further restrictions were placed on participant populations, nor were restrictions placed on publication status.

After executing the automated search procedure, two authors conducted the manual search. Five authors independently screened the resulting titles and abstracts against the

---

[1] Public calls were made on Twitter (these tweets were viewed approximately 10,000 times) and the Psychological Methods Discussion Group on Facebook.

inclusion criteria. The full texts of included studies were then retrieved and screened. Any disagreements were discussed and, if needed, an additional author was consulted. Finally, two authors conducted forward and backward reference-list searches from the included studies.

Data Extraction and Management

Relevant data were extracted from eligible studies and entered into a spreadsheet. Elements extracted include: publication year, a description of the study population involved, study sample size, the source of logged and self-reported data, the form of media use recorded, and measurement produced (e.g., total use, average use, etc.), and the duration for which logged data was acquired. To enable the analysis of convergent validity, effect sizes were extracted from reported correlation analyses for associations between self-reported and logged measures of media use as well as for correlations between problematic use and logged measures. For estimates of use, we only included comparisons for equivalent actions, time periods, and forms (e.g., average phone use per day, total weekly social media use, or daily phone pickups etc.) while, for problematic use scales, we included reported associations with logged measures for the duration or volume of use for any of the five targeted media (e.g., total phone time, average phone pickups, etc.). Both Pearson's product moment correlation coefficients (r) and Spearman's rank-ordered correlation coefficients (r_s) were extracted.

To analyse under- or over-reporting, we extracted measures of central tendency and variability for self-reported estimates that explicitly concern either the duration or the volume of media use reported on a continuous scale and logged measures for equivalent outcomes. To perform moderator analyses, we coded media as either 'phone', 'gaming', 'social media', 'computer', or 'Internet'. This categorisation was based on the source of log-tracked data and, in instances in which overlap existed (e.g., social media on a phone), we coded the most specific medium known. Self-report measures were coded to capture one of two outcomes: ‘use’ or ‘problematic use’, reflect one of two forms: ‘scale’ or ‘single estimate’, and represent one of two categories of use: ‘duration’ or ‘volume’ (i.e., use instances).

When reported data were insufficient to compute the necessary effect sizes, we contacted the corresponding authors to request ad hoc analyses or for further descriptive statistics. If, after two attempts the relevant data were still not available, and relevant values were represented in plots in a paper, we used a web plot digitizer (Rohatgi, 2019) to convert plotted representations into numerical values. If no response was received from corresponding authors and relevant plots were not available to be digitized, the comparison was excluded.

Data Analysis

All analyses were performed with the R statistical programming language (v. 4.0.2; R Core Team, 2017). A complete list of the packages used in the analysis is provided in the analysis code available through the OSF.[2] Three distinct meta-analyses were conducted. In the first, we focused on correlations between self-reported and logged media use. In the second, the analysis concerned the degree of under- or over-reporting. In the third, we focused on correlations between self-reported *problematic* media use and logged media use. For all analyses we adopted an a priori statistical significance level of $\alpha = .05$. To account for variance inflation resulting from dependent observations for different measures for the same participants, we used cluster-robust variance estimation (RVE) based on the sandwich method with adjusted estimators

[2] Deviating from the protocol, the robust variance estimation was conducted with the *robumeta* package rather than the *metafor* package as specified.

for small samples and a correlated effects weighting scheme using the default assumed value of r = 0.80 (Hedges et al., 2010). For all moderator analyses, acknowledging that there is no widely accepted minimum number of effects required, noting Fu et al. (2011)'s recommendation, we specified a minimum requirement of four included effects per moderator level.

For the correlational meta-analyses, to stabilise the variances, raw effect sizes were transformed into normalised correlation coefficients (Fisher's z). Effects reported as Spearman's r_s were first transformed to Pearson's r, as Gilpin (1993) describes, and then transformed to Fisher's z for synthesis. For reporting, we performed Fisher's z-to-r transformation (Borenstein et al., 2011). For both correlational meta-analyses, we estimated random-effects models to calculate overall summary effect sizes.

To analyse usage correlations the analysis only included effect sizes for correlations between logged usage and self-report measures that explicitly concerned media use. For these analyses, if a study reported correlations for both logged overall use (total or average duration or volume) and logged use of specific smartphone applications or websites, to avoid nested correlations, we excluded correlations involving individual applications or websites and only included comparisons for overall indications of use. However, if an otherwise eligible comparison was reported and no overall use metric was available, comparisons for specific use types were included. Furthermore, if no comparison with overall use was reported, with the exception of social media and gaming, we excluded comparisons that involved aggregations of different applications or websites into higher-level categories (i.e., use of navigation applications, use of video platforms, use of fitness applications etc.). To analyse correlations for measures concerning problematic media use, the analysis only included effect sizes for correlations between logged media use and self-reported problematic media use.

To interpret the outcomes of the correlational meta-analyses, in-line with Cohen (1988), we took correlation coefficients of .1 to be small, .30 to be medium, and .50 or greater to be large effect sizes, respectively. However, noting our aim of investigating convergent validity, acknowledging Carlson and Herdman (2012)'s recommendations, we considered correlation coefficients above 0.7 to indicate strong evidence of convergent validity, between 0.5 and 0.7 to indicate acceptable convergent validity, and below 0.5 to be inadequate to support convergent validity between the two measurement forms.

To consider possible factors that affect the relationship between self-reported and logged media use, three categorical moderator analyses were conducted. The first concerned the effect of the medium on the association between self-reported and log-based media use. The second considered the potential moderating effect of the measure category (either volume or duration), while the third concerned the form of self-report measure (scale or estimate). For each moderator category, we estimated separate random effects models to produce summary weighted effect sizes for each subgroup.

To investigate measurement accuracy, we only considered single point estimates for overall use duration or use instances for a given medium that were provided on a continuous scale. We first determined the proportion of comparisons that are indicative of accurate, under-reported, or over-reported media use. For this analysis, we used a margin of error of 5% or more above the tracked measure to indicate over-reporting, 5% or more below to indicate under-reporting, and mean estimates within 5% of the logged measure to be accurate. To quantify the magnitude of the difference in means produced using the different measurement forms, given the within-subjects nature of the analysis and the existence of a true ratio scale with a natural zero point (Borenstein et al., 2011), we calculated the log transformed ratio of means (Friedrich et al., 2008; Lajeunesse, 2011), and estimated the sampling variance accounting for the correlation between measurements (Viechtbauer, 2010). These unitless effect sizes were then synthesized by estimating a random effects model and then back transformed for reporting.[3] In this analysis, a value of one corresponds to an equal ratio between self-reported and logged measures, while values less than one indicate under-reporting and values greater than one indicate over-reporting. The magnitude of the outcome represents the ratio of self-reported to logged media use. We planned two categorical moderator analyses, estimating random effects models to produce summary weighted effect sizes for each subgroup. In the first, we examined whether the results differed based on the category of use estimated (e.g., use duration or use volume). In the second, we examined whether they differ by the medium.

For the three primary meta-analyses, to examine the variance and heterogeneity among effects, we computed Q and I^2, interpreting statistically significant Q values to indicate heterogeneity and I^2 values of approximately 25%, 50%, and 75% to indicate low, moderate, and high heterogeneity, respectively. To determine if the analyses were impacted by any outliers, we conducted outlier and influence diagnostics for the original models (i.e., Cook's distance, covariance ratios, diagonal elements of the hat matrix; Viechtbauer, 2010), and performed sensitivity re-analyses without any outliers. Equivalence testing using the two one-sided test (TOST) procedure was also applied to assess evidence for the absence of meaningful effects. A smallest effect size of interest of $r = 0.1$ was used to determine equivalence bounds (i.e., a lower bound of -0.1 and a higher bound of 0.1). The results of the TOST procedure are presented in the Supplementary Information.

---

[3] This ratio of means is commonly known as the response ratio (R) in Ecology research.

To account for study quality and assess potential biases due to ‘small-study effects’, which can include publication bias, we visually inspected funnel plot symmetry and performed Egger’s regression test (Egger et al., 1997). To visualize possible publication bias, we used a contour-enhanced funnel plot which superimposes notable areas of statistical significance (i.e., $p = 0.1$, $p = 0.05$, $p = 0.01$). An over-representation of effect sizes in the highlighted areas is indicative of possible publication biases (Peters et al., 2008). As a further sensitivity analysis, we conducted moderator analyses to determine if effect sizes reported in peer-reviewed studies differ from pre-publication studies (e.g., preprints, unpublished data, or papers under review).

References

Abendroth, A., Parry, D. A., le Roux, D. B., & Gundlach, J. (2020). An analysis of problematic media use and technology use addiction scales–what are they actually assessing? In *Conference on e-Business, e-Services and e-Society* (pp. 211-222). Springer, Cham.

Ahn, H., Nguyen, G. T., Lee, H., Jo, S. J., Kim, E. J., & Yim, H. W. (2017). Data collection and analysis of smartphone use and sleep of secondary school children. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 410-413). IEEE.

American Academy of Pediatrics. (2016). Media and young minds. Pediatrics, 138(5).

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PloS one*, 10(10).

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of Internet use. *Communication Methods and Measures*, *11*(3), 173-190.

Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. *Journal of Computer-Mediated Communication,* 18(4), 508–519.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.

Brasel, S. A., & Gips, J. (2017). Media multitasking: How visual cues affect switching behavior. *Computers in Human Behavior*, 77, 258–265.

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1909-1912).

Burnell, K., Ackerman, R. A., George, M., J., Kurup, A. R., & Underwood, M. K., (unpublished). Unpublished data received directly from the authors.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods,* 15(1), 17–32.

Chuck, E. (2017). Is social media contributing to rising teen suicide rate?. Available at https://nbcnews.com/news/us-news/social-media-contributing- rising-teen-suicide-rate-n812426 (Retrieved September 21, 2020).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Academic press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. P*sychological Bulletin,* 52(4), 281.

Davidson, B. I., Ellis, D. A., Bowman, P., Nicholas D, Liveley, G., Shaw, H., Przybylski, A. K., & Levine, M. (2020a). Avoiding irrelevance: The manifestation and impacts of technophobia in psychological science. PsyArXiv.

Davidson, B. I., Shaw, H., & Ellis, D. A. (2020b). Fuzzy constructs in assessment: The overlap between mental health and technology 'use'. PsyArXiv.

Deng, T., Kanthawala, S., Meng, J., Peng, W., Kononova, A., Hao, Q., ... & David, P. (2019). Measuring smartphone usage and task switching with log tracking and self-reports. *Mobile Media & Communication*, *7*(1), 3-23.

Dickson, K., Richardson, M., Kwan, I., Macdowall, W., Burchett, H., Stansfield, C., & Thomas, J. (2019). Screen-based activities and children and young people's mental health and psychosocial wellbeing: a systematic map of reviews. *London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.*

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ: British Medical Journal,* 315(7109), 629–634.

Elhai, J. D., Tiamiyu, M. F., Weeks, J. W., Levine, J. C., Picard, K. J., & Hall, B. J. (2018). Depression and emotion regulation predict objective smartphone use measured over one week. *Personality and Individual Differences*, *133*, 21-28.

Ellis, D. A. (2019). Are smartphones really that bad? improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior,* 97, 60–66.

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? *International Journal of Human-Computer Studies,* 130, 86–92.

Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Felisoni, D. D., & Godoi, A. S. (2018). Cell phone usage and academic performance: An experiment. *Computers & Education*, *117*, 175-187.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychology*

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science,* 8(4), 370–378.

Friedrich, J. O., Adhikari, N. K., & Beyene, J. (2008). The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. *BMC Medical Research Methodology*, 8(1), 32.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., ... & Santaguida, P. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of clinical epidemiology*, 64(11), 1187-1197.

Gerpott, T. J., & Thomas, S. (2014). Empirical research on mobile internet usage: A meta-analysis of the literature. *Telecommunications Policy,* 38(3), 291–310.

Geyer, K., Carbonell, X., Beranuy, M., Calvo, F, & Ellis, D. A. (unpublished). Unpublished data received directly from the authors.

Gilpin, A. R. (1993). Table for conversion of kendall's tau to spearman's rho within the context of measures of magnitude of effect for meta-analysis. *Educational and psychological measurement,* 53(1), 87–92.

Griffioen, N., Rooij, M. van, Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward Improved Methods in Social Media Research. Technology, Mind, and Behavior, 1(1).

Harris, B., Regan, T., Schueler, J., & Fields, S. A. (2020). Problematic mobile phone and smartphone use scales: A systematic review. *Frontiers in Psychology*, 11, 672.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, 1(1), 39–65.

Howard, M. C., & Jayne, B. S. (2015). An analysis of more than 1,400 articles, 900 scales, and 17 years of research: The state of scales in cyberpsychology, behavior, and social networking. *Cyberpsychology, Behavior, and Social Networking*, 18(3), 181–187.

Inyang, I., Benke, G., Morrissey, J., McKenzie, R., & Abramson, M. (2009). How well do adolescents recall use of mobile telephones? Results of a validation study. *BMC Medical Research Methodology*, *9*(1), 36.

Jobe, J. B. (2003). Cognitive psychology and self-reports: models and methods. *Quality of Life Research*, 12(3), 219–227.

John, O. P., & Benet-Martíınez, V. (2014). *Measurement: Reliability, construct validation, and scale construction.* In H. T. Reis & C. M. Judd (Eds.), Handbook of research methods in social and personality psychology (p. 473–503). Cambridge, UK: Cambridge University Press.

Jones-Jang, S. M., Heo, Y. J., McKeever, R., Kim, J. H., Moscowitz, L., & Moscowitz, D. (2020). Good News! Communication Findings May be Underestimated: Comparing Effect Sizes with Self-Reported and Logged Smartphone Use Data. *Journal of Computer-Mediated Communication*.

Junco, R. (2013). Comparing actual and self-reported measures of facebook use. *Computers in Human Behavior,* 29(3), 626–631.

Jungselius, B., & Weilenmann, A. (2018). Conceptualizing 'use' in social media studies. In *Proceedings of the 9th international conference on social media and society* (pp. 325–329).

Jürgens, P., Stark, B., & Magin, M. (2019). Two half-truths make a whole? on bias in self-reports and tracking data. *Social Science Computer Review,* 0894439319831643.

Kahn, A. S., Ratan, R., & Williams, D. (2014). Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication,* 19(4), 1010–1023.

Katapally, T. R., & Chu, L. M. (2019). Methodology to derive objective screen-state from smartphones: A SMART platform study. *International journal of environmental research and public health*, *16*(13), 2275.

Kaye, L. K., Orben, A., Ellis, D. A., Hunter, S. C., & Houghton, S. (2020). The Conceptual and Methodological Mayhem of "Screen Time." *International Journal of Environmental Research and Public Health*, 17(10), 3661.

Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use. *Communication Methods and Measures*, *6*(2), 126-143.

Laconi, S., Rodgers, R. F., & Chabrol, H. (2014). The measurement of internet addiction: A critical review of existing scales and their psychometric properties. *Computers in Human Behavior*, 41, 190–202.

Lajeunesse, M. J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11), 2049–2055.

Lee, H., Ahn, H., Choi, S., & Choi, W. (2014). The SAMS: Smartphone addiction management system and verification. *Journal of medical systems*, *38*(1), 1.

Lee, H., Ahn, H., Nguyen, T. G., Choi, S. W., & Kim, D. J. (2017). Comparing the self-report and measured smartphone usage of college students: a pilot study. *Psychiatry investigation*, *14*(2), 198.

Lin, Y. H., Lin, Y. C., Lee, Y. H., Lin, P. H., Lin, S. H., Chang, L. R., ... & Kuo, T. B. (2015). Time distortion associated with smartphone addiction: Identifying smartphone addiction via a mobile application (App). *Journal of psychiatric research*, *65*, 139-145.

Loid, K., Täht, K., & Rozgonjuk, D. (2020). Do pop-up notifications regarding smartphone use decrease screen time, phone checking behavior, and self-reported problematic smartphone use? Evidence from a two-month experimental study. *Computers in Human Behavior*, *102*, 22-30.

Malinauskas, R., & Malinauskiene, V. (2010). A meta-analysis of psychological interventions for Internet/smartphone addiction among adolescents. Journal of Behavioral Addictions, 8(4), 613–624. https://doi.org/10.1556/2006.8.2019.72

Meier, A., & Reinecke, L. (2020, April 6). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. https://doi.org/10.31234/osf.io/573ph

Mikkelsen, S., Vilstrup, I., Lassen, C. F., Kryger, A. I., Thomsen, J. F., & Andersen, J. H. (2007). Validity of questionnaire self-reports on computer, mouse and keyboard usage during a four-week period. *Occupational and Environmental Medicine*, *64*(8), 541-547.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, 6(7).

Montag, C., Błaszkiewicz, K., Lachmann, B., Sariyska, R., Andone, I., Trendafilov, B., & Markowetz, A. (2015). Recorded behavior as a valuable resource for diagnostics in mobile phone addiction: evidence from psychoinformatics. *Behavioral Sciences*, *5*(4), 434-442.

Newell, C. A. (2018). *Smartphone and Social Media use and its Health Associations* (Masters dissertation). The University of Texas at Arlington.

Newton, C. (2019). New legislation is putting social networks in the crosshairs. The Verge. Retrieved at https://www.theverge.com/2019/8/1/20749517/social-network-legislation-hawley-privacy-research (Retrieved September 27, 2020).

Noë, B., Turner, L. D., Linden, D. E., Allen, S. M., Winkens, B., & Whitaker, R. M. (2019). Identifying indicators of smartphone addiction through user-app interaction. *Computers in human behavior*, *99*, 56-65.

Nunnally, J. (1994). *Psychometric theory 3rd Edition*. Tata McGraw-Hill Education.

Oeldorf-Hirsch, A., & Chen, Y. (2020). Who Cares about Screen Time? Predicting the use of Mobile Phone Tracking Features. Paper presented at *the ICA Annual Conference 2020*, Virtual (unpublished).

Ohme, J., Araujo, T., de Vreese, C., Piotrowski, J. (2020). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. Paper presented at *the ICA Annual Conference 2020*, Virtual (unpublished).

Orben, A. (2020). The Sisyphean Cycle of Technology Panics. *Perspectives on Psychological Science*, *15*(5), 1143–1157.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour,* 3(2), 173.

Pan, Y. C., Lin, H. H., Chiu, Y. C., Lin, S. H., & Lin, Y. H. (2019). Temporal stability of smartphone use data: Determining fundamental time unit and independent cycle. *JMIR mHealth and uHealth*, *7*(3).

Parry, D., Davidson, B. I., Sewall, C., Fisher, J. T., Mieczkowsk*i,* H., & Quintana, D. S. (2020). Associations Between Log-Based and Self-Reported Digital Media Use: A Systematic Review and Meta-analysis Protocol. Retrieved at https://doi.org/10.17605/OSF.IO/KCSHV (Retrieved September 21, 2020).

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of clinical epidemiology*, 61(10), 991–996.

Piwek, L., Ellis, D. A., & Sally, A. (2016). Can programming frameworks bring smartphones into the mainstream of psychological science? *Frontiers in psychology,* 7, 1252.

Prasad, S., Harshe, D., Kaur, N., Jangannavar, S., Srivastava, A., Achanta, U., ... & Harshe, G. (2018). A study of magnitude and psychological correlates of smartphone use in medical students: A pilot study with a novel telemetric approach. *Indian journal of psychological medicine*, *40*(5), 468-475.

Quintana, D. S. (2015). From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in psychology*, 6, 1549.

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria. R Foundation for Statistical Computing

Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological methods*, 10.1037/met0000300.

Rohatgi, A., 2019. WebPlotDigitizer. Retrieved at: https://apps.automeris.io/wpd/ (Retrieved September 21, 2020).

Rosen, L. D., Carrier, L. M., Pedroza, J. A., Elias, S., O'Brien, K. M., Lozano, J., ... & Ruiz, A. (2017). The role of executive functioning and technological anxiety (FOMO) in college course performance as mediated by technology usage and multitasking habits. *Educational Psychology*, *24*(1), 14-25.

Rozgonjuk, D., Levine, J. C., Hall, B. J., & Elhai, J. D. (2018). The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Computers in Human Behavior*, *87*, 10-17.

Rozgonjuk, D., Pruunsild, P., Jürimäe, K., Schwarz, R. J., & Aru, J. (2020). Instagram use frequency is associated with problematic smartphone use, but not with depression and anxiety symptom severity. *Mobile Media & Communication*, 2050157920910190.

Ryding, F. C., & Kuss, D. J. (2020). Passive objective measures in the assessment of problematic smartphone use: A systematic review. *Addictive Behaviors Reports*, 100257.

Scharkow, M. (2016). The accuracy of self-reported internet use—a validation study using client log data. *Communication Methods and Measures*, 10(1), 13–27.

Sela, Y., Zach, M., Amichay-Hamburger, Y., Mishali, M., & Omer, H. (2020). Family environment and problematic internet use among adolescents: the mediating roles of depression and fear of missing out. *Computers in Human Behavior*, *106*, 106226.

Sen, I., Floeck, F., Weller, K., Weiss, B., & Wagner, C. (2019). A total error framework for digital traces of humans. *arXiv preprint arXiv:1907.08228*.

Sewall, C. J., Bear, T. M., Merranko, J., & Rosen, D. (2020). How psychosocial well-being and usage amount predict inaccuracies in retrospective estimates of digital technology use. *Mobile Media & Communication*.

Shaw, H., Ellis, D., Geyer, K., Davidson, B. I., Ziegler, F., & Smith, A. (2020). Subjective reports overstate the relationship between screen time and mental health.

Shin, C., & Dey, A. K. (2013). Automatically detecting problematic use of smartphones. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 335-344).

Shin, M., & Lee, K. (2017). Measuring smartphone usage time is not sufficient to predict smartphone addiction. *Journal of Theoretical and Applied Information Technology*, *95*(19).

Shum, M., Kelsh, M. A., Sheppard, A. R., & Zhao, K. (2011). An evaluation of self-reported mobile phone use compared to billing records among a group of engineers and scientists. *Bioelectromagnetics*, *32*(1), 37-48.

Singh, V. K., & Jain, A. (2017). Toward harmonizing self-reported and logged social data for understanding human behavior. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2233-2238).

Social Media Addiction Reduction Technology Act, S. 2314, 116th cong. (2019). https://www.congress.gov/bill/116th-congress/senate-bill/2314/all-info

Timotijevic, L., Barnett, J., Shepherd, R., & Senior, V. (2009). Factors influencing self-report of mobile phone use: the role of response prompt, time reference and mobile phone use in recall. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(5), 664-683.

Tokola, K., Kurttio, P., Salminen, T., & Auvinen, A. (2008). Reducing overestimation in reported mobile phone use associated with epidemiological studies. *Bioelectromagnetics: Journal of the Bioelectromagnetics Society, The Society for Physical Regulation in Biology and Medicine, The European Bioelectromagnetics Association*, *29*(7), 559-563.

Tourangeau, R., et al. (1984). Cognitive sciences and survey methods. Cognitive aspects of survey methodology: Building a bridge between disciplines, 15, 73–100.

Twenge, J. M. (2017). Have smartphones destroyed a generation. *The Atlantic*. Retrieved at: https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/ (Retrieved May 19, 2020) van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies*, *125*, 118-128.

Vanden Abeele, M., Beullens, K., & Roe, K. (2013). Measuring mobile phone use: Gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use. *Mobile Media & Communication*, 1(2), 213–236.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software,* 36(3), 1–48.

Vorderer, P., Krömer, N., & Schneider, F. M. (2016). Permanently online–permanently connected: Explorations into university students' use of social media and mobile smart devices. *Computers in Human Behavior*, 63, 694–703.

Webster, J., & Watson, R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2).

Wilmer, H. H., Hampton, W. H., Olino, T. M., Olson, I. R., & Chein, J. M. (2019). Wired to be connected? Links between mobile technology engagement, intertemporal preference and frontostriatal white matter connectivity. *Social cognitive and affective neuroscience*, *14*(4), 367-379.

Yeykelis, L., Cummings, J. J., & Reeves, B. (2014). Multitasking on a single device: arousal and the frequency, anticipation, and prediction of switching between media content on a computer. *Journal of Communication,* 64(1), 167–192.

Yuan, N., Weeks, H. M., Ball, R., Newman, M. W., Chang, Y. J., & Radesky, J. S. (2019). How much do parents actually use their smartphones? Pilot study comparing self-report to passive sensing. *Pediatric research*, *86*(4), 416-418.

Supplementary Information

Supplementary Table 1. Overview of included studies for the self-reported media use meta-analyses.

Authors and year	k	n	Medium	Self-Report measure	Logged measure
Ahn et al. (2017)	1	23	Phone	Estimated daily usage duration [a]	SAMS Monitor
Andrews et al. (2015)	2	23	Phone	Estimated daily usage duration [a]	Custom application
		23	Phone	Estimated daily pickups [a]	Custom application
Araujo et al. (2017)	1	690	Internet	Estimated daily usage duration [a]	Internet use tracking software
Boase and Ling (2013)	2	426	Phone	Estimated call number for previous day [a]	Telecom provider log data
		426	Phone	Estimated text number for previous day [a]	Telecom provider log data
Burke et al. (2010)	1	155	Soc Med	Usage item from the FIS	Facebook server logs
Burnell et al. (unpublished)[c]	1	1584	Soc Med	Estimated daily usage time (Facebook, Instagram, Twitter, SnapChat) [a]	iOS Screen Time
Deng et al. (2018)	1	44	Phone	Estimate of app usage time [a]	App Usage Tracker
Ellis et al. (2019	3	238	Phone	Estimated daily usage duration [a]	iOS Screen Time
		238	Phone	Estimated daily pickups [a]	iOS Screen Time
		238	Phone	Daily smartphone usage subscale of the MTUAS	iOS Screen Time
Ernala et al. (2020)	6	5000	Soc Med	Estimated daily Facebook usage duration	Facebook log data
		5000	Soc Med	Estimated daily Facebook usage duration in past week	Facebook log data
		5000	Soc Med	Estimated daily Facebook usage duration	Facebook log data
		5000	Soc Med	Estimated daily Facebook usage duration	Facebook log data
		5000	Soc Med	Estimated daily Facebook checks	Facebook log data
		5000	Soc Med	Estimated daily Facebook checks	Facebook log data
Felisoni and Godoi (2017)	1	43	Phone	Estimated daily usage duration [a]	Moment/App Usage Tracker

Geyer et al. (unpublished)[c]	1	139	Phone	Estimated daily usage duration [a]	Activity Logger
Inyang et al. (2009)	2	59	Phone	Estimated weekly call duration [a, b]	Phone use logger
		59	Phone	Estimated weekly call number [a]	Phone use logger
Jones-Jang et al. (2020)	4	294	Phone	Estimated daily usage duration [a]	iOS Screen Time
		294	Phone	Estimated daily pickups [a]	iOS Screen Time
		291	Phone	Estimated daily usage duration [a]	iOS Screen Time
		291	Phone	Estimated daily pickups [a]	iOS Screen Time
Junco (2013)	2	45	Soc Med	Estimated daily Facebook usage duration [a]	Custom application
		45	Soc Med	Estimated daily Twitter usage duration [a]	Custom application
Kahn (2014)	1	6598	Games	Estimated weekly usage duration [a]	Platform logged data
Katapally and Chu (2019)	1	49	Phone	Usage duration items from modified sedentary behaviour scale [b]	Screen-state sensor
Kobayashi and Boase (2012)	4	270	Phone	Estimated outgoing call number [a]	Custom application
		270	Phone	Estimated incoming call number [a]	Custom application
		182	Phone	Estimated outgoing text number [a]	Custom application
		182	Phone	Estimated incoming text number [a]	Custom application
Lee et al. (2017)	1	35	Phone	Estimated weekly usage duration [a, b]	Custom application
Lin et al. (2015)	1	66	Phone	Assisted estimate of total usage duration [a]	Custom application
Mikkelsen et al. (2007)	1	1211	Computer	Estimated weekly usage duration [a]	WorkPace software
Montag et al. (2015)	1	58	Phone	Estimated weekly usage duration [a]	Custom application
Newell et al. (2018) [c]	1	135	Soc Med	Usage item from the FIS	iOS battery logs/Android PhoneUsage
Oeldorf-Hirsh and Chen (unpublished) [c]	1	142	Phone	Estimated daily usage duration [a]	iOS Screen Time
Ohme et al. (unpublished) [c]	2	47	Phone	Estimated daily usage duration [a]	iOS Screen Time
		47	Phone	Estimated daily pickups [a]	iOS Screen Time
Rosen et al. (2018)	2	216	Phone	Daily smartphone usage subscale of the MTUAS	Instant Quantified Self
		104	Phone	Daily smartphone usage subscale of the MTUAS	Instant Quantified Self
Rozgonjuk et al. (2020)	2	30	Phone	Estimated daily Instagram usage duration [a]	App Usage Manage/Track Usage
		30	Phone	Estimated daily Instagram checks [a]	App Usage Manage/Track

					Usage
Scharkow (2016)	1	3401	Internet	Estimated weekly private usage duration	Browser logs
Sewall et al. (2020)	2	325	Phone	Estimated weekly usage duration [a]	iOS Screen Time
		325	Phone	Estimated daily usage duration [a]	iOS Screen Time
Shaw et al. (2020) [c]	4	46	Phone	Estimated daily usage duration [a]	Activity Logger
		46	Phone	Estimated daily pickups [a]	Activity Logger
		199	Phone	Estimated daily usage duration [a]	iOS Screen Time
		199	Phone	Estimated daily pickups [a]	iOS Screen Time
Shum et al. (2011)	2	60	Phone	Estimated total call duration [a]	Billing records
		60	Phone	Estimated number of calls [a]	Billing records
Singh and Jain (2017)	1	47	Phone	Estimated number of calls	Call meta-data
Timotijevic et al. (2009)	2	159	Phone	Estimated outgoing number of calls [a]	Billing records
		173	Phone	Estimated outgoing call duration [a]	Billing records
Tokala et al. (2008)	1	70	Phone	Estimated monthly call duration	Telecom provider log data
van Berkel et al. (2018)	1	20	Phone	Estimated daily usage duration [a]	Custom application
Vanden Abeele et al. (2013)	3	466	Phone	Estimated weekly call number [a]	Telecom provider log data
		466	Phone	Estimated weekly call duration [a]	Telecom provider log data
		466	Phone	Estimated weekly text number [a]	Telecom provider log data
Wilmer et al. (2019)	1	56	Phone	Mobile technology engagement scale	iOS battery logs
Yuan et al. (2019)	2	38	Phone	Estimated usage duration yesterday	Moment/Minuku
		50	Phone	Estimated daily pickups	Moment/Minuku

Note: k: number of separate effect sizes included from a paper (many papers include separate studies with distinct samples); n: sample size. For self-report measure: FIS = Facebook intensity scale; MTUAS = Media and technology usage and attitudes scale.
a: Included in analysis of under- or over-reporting.
b: Digitiser used to retrieve complete data.
c: Non-peer reviewed at the time of inclusion.

Supplementary Table 2. Overview of included studies for the self-reported problematic media use meta-analysis.

Authors and year	k	n	Medium	Self-Report measure	Logged measure
Andrews et al. (2015	2	23	Phone	MPPUS	Custom application
		23	Phone	MPPUS	Custom application
Elhai et al. (2018)	1	68	Phone	SAS-SV	Moment
Ellis et al. (2019)	12	238	Phone	MPPUS	iOS Screen Time
		238	Phone	MPPUS	iOS Screen Time
		238	Phone	NQ	iOS Screen Time
		238	Phone	NQ	iOS Screen Time
		238	Phone	SAS	iOS Screen Time
		238	Phone	SAS	iOS Screen Time
		238	Phone	SABAS	iOS Screen Time
		238	Phone	SABAS	iOS Screen Time
		238	Phone	PMPUQ	iOS Screen Time
		238	Phone	PMPUQ	iOS Screen Time
		238	Phone	SUQ-A	iOS Screen Time
		238	Phone	SUQ-A	iOS Screen Time
Geyer et al. (unpublished) [a]	1	139	Phone	SAS-SV	Usage Logger
Jones-Jang et al. (2020)	2	294	Phone	SAS	iOS Screen Time
		291	Phone	SAS	iOS Screen Time
Lee et al. (2014)	2	14	Phone	K-SAS	Custom application
		14	Phone	K-SAS	Custom application
Loid et al. (2020)	2	45	Phone	ESAPS-SV	App Usage manager
		45	Phone	ESAPS-SV	App Usage manager
Montag et al. (2015)	1	58	Phone	MPPUS	Custom application
Noë et al. (2019)	1	64	Phone	SAS	Tymer
Pan et al. (2019	2	33	Phone	SPAI-5	Smartphone use logger
		33	Phone	SPAI-5	Smartphone use logger
Prasad et al. (2018)	3	140	Phone	SAS	App Usage tracker

		140	Phone	SAS	Instant Quantified Self
		140	Phone	SAS	Instant Quantified Self
Rozgonjuk et al. (2018)	2	101	Phone	SAS	Moment
		101	Phone	SAS	Moment
Sela et al. (85)	1	85	Internet	GPIUS	Mobile online activity logger
Shaw et al. (2020) [a]	4	46	Phone	SAS	Activity Logger
		46	Phone	SAS	Activity Logger
		199	Phone	SAS	iOS Screen Time
		199	Phone	SAS	iOS Screen Time
Shin and Dey (2013)	1	48	Phone	Modified MPPUS	Custom application
Shin and Lee (2017)	1	195	Phone	Modified SASDS	Smartphone Usage Tracker
Wilmer et al. (2019)	2	56	Phone	MMI	iOS battery logs
		56	Phone	MPPUS	iOS battery logs

Note: k: number of separate effect sizes included from a paper (many papers include separate studies with distinct samples); n: sample size. For self-report measure, MPPUS = mobile phone problem use scale; SAS = smartphone addiction scale (SV = short version; K = Korean version); NQ = nomophobia questionnaire; SABAS = smartphone application-based addiction scale; PMPUQ = problematic mobile phone use questionnaire; SUQ-A = smartphone use questionnaire-absent minded; ESAPS-SV = Estonian smartphone addiction proneness scale-short version; SPAI-5 = smartphone addiction inventory; GPIUS = generalised problematic internet use scale; SASDS = smartphone addiction self-diagnosis scale; MMI = media multitasking inventory.
a: Non-peer reviewed at the time of inclusion.

Supplementary Analyses

For associations between self-reported and logged media use a TOST analysis confirmed that the observed effect is statistically different from zero and statistically not equivalent to zero, given equivalence bounds of $r = -0.1$ to 0.1 ($Z = 10.89$, $p = 1.00$). Similarly, for associations between self-report measures of problematic media use and logged media use a TOST analysis, the observed effect is statistically different from zero and statistically not equivalent to zero, given equivalence bounds of $r = -0.1$ to 0.1 ($Z = 5.62$, $p = 1.000$). Notably, the equivalence bounds of $r = -0.1$ and 0.1 were converted to $d = -0.201$ and 0.201 for compatibility.

Following communication delays, a single study (Yuan et al., 2019) was included in the analysis after the publication of a preprint containing the preliminary results of the meta-analyses. As a robustness check, we considered the summary effect size for usage correlations with ($r = 0.38$, 95% CI [0.33, 0.42], $p < 0.001$) and without ($r = 0.38$, 95% CI [0.33, 0.42], $p < 0.001$) the two additional effect sizes included from this study. As is evident, the nature and magnitude of the summary effect size was not impacted by the inclusion of these additional effects.