Does working from home improve employees' productivity? Empirical evidence from a meta-analysis

Colin F. Mang*

Department of Economics, McMaster University

Amar Anwar

Shannon School of Business, Cape Breton University

International SEPT Competence Center, Universität Leipzig

*Corresponding Author: Colin F. Mang, Department of Economics, McMaster University, 1280 Main St. W., Hamilton, ON, Canada, L8S 4L8, Email: **colinmang@mcmaster.ca**

Working From Home (WFH) and Productivity

- Productivity enhancing effects:
  - Greater motivation due to autonomy (Bailey and Kurland, 2002)
  - Reduces stress and energy loss from commuting and also frees up more time to devote to work (Kanellopoulos, 2011; Kazekami, 2020)
  - Provides greater work-life balance, especially valuable for parents (pre-pandemic) (Arntz et al., 2022)

- Productivity diminishing effects:
  - Lack of monitoring leads to shirking (Causer and Jones, 1996; Day and Burbach, 2015; Pyöriä, 2011)
  - Personal life disruptions (Mann and Holdsworth, 2003; Delanoeije et al., 2019)
  - Reduced communications, weaker team building, and lower collaboration (Baruch, 2000; Fonner and Roloff, 2010; Sarker et al., 2012)

Empirical Evidence

- Productivity Increase:
  - Bergeaud et al. (2023a), Bergeaud et al. (2023b), Chen et al. (2023); Labrado Antolín et al. (2024); Lee and Gascó-Hernandez (2023); Neirotti et al. (2012); Peters et al. (2008)
- Productivity Decrease:
  - Giménez-Nadal et al. (2020); Kitigawa et al. (2021a); van der Lippe and Lippényi (2019)
- Mixed Results:
  - Arntz et al. (2022), Baek (2021); Choudhury et al. (2021); Deole et al. (2023); Kazekami (2020); Monteiro et al. (2021); Morikawa (2022)
- Insignificant Impact:
  - Kawaguchi and Motegi (2021)

Research Questions

1. Does WFH increase productivity?

2. Did the pandemic WFH experience differ from the pre-pandemic WFH experience?

3. Does WFH productivity vary by sex, parental status, and sector of employment?

4. Can differences in study methodologies explain the conflicting findings in the literature?

The Literature

- We examine 59 studies that provide 2,688 estimates of Equation (1):

$$Productivity_i = \alpha WFH_i + \varphi X_i + \varepsilon_i \quad (1)$$

where X are other factors affecting productivity such as gender, parental status, worker education, firm size, commuting distance, etc. and ε is an error term; variables contained in X vary from study to study.

- α is the parameter of interest: the impact of working from home on employee productivity. We collect $\hat{\alpha}$ from the studies.

- We separate studies using pre-pandemic data from studies using data collected during the pandemic due to differences in WFH participation, working conditions, and technology

Differing Methodologies

- Is participation in a WFH program exogenously determined or endogenous?
  - Endogenous: Kawaguchi and Motegi (2021); Lee and Gascó-Hernandez (2023); Neirotti et al. (2012); Pigini and Staffolani (2019); Schroeder and Warren (2005).

- Data collected from:
  - Employer reports (e.g., Abrardi et al., 2022; Monteiro et al., 2021; Neirotti et al., 2012).
  - Employee self-reports (e.g., Deole et al., 2023; Kazekami, 2020).

- Data collection method:
  - Single cross-section (e.g., Kitigawa et al., 2021a; Pabilonia and Vernon, 2022).
  - Panel or Pseudo-panel (e.g., Arntz et al., 2022; Braun et al., 2021).

Differing Methodologies

- Measures of productivity:
  - Income earned (e.g., Kazekami, 2020; Kouki and Sauer, 2022).
  - Physical output (e.g., Baek, 2021; Choudhury et al., 2021).
  - Revenue generated (e.g., Kawaguchi et al., 2022); Monteiro et al., 2021).
  - Total hours worked (e.g., Lyttelton et al., 2020; Peters et al., 2008).
  - Likert scale (e.g., Abrardi et al., 2022; Lee and Gascó-Hernandez, 2023).

- Measures of WFH:
  - Binary (Yes/No) (e.g., Arntz et al., 2022; Deole et al., 2023).
  - Continuous (Time spent WFH) (e.g., Labrado Antolín et al., 2024; Zarifhonarvar et al., 2023).

Standardized *PCC*

- The Partial Correlation Coefficient (PCC) standardizes the estimates of α from Equation (1) to account for the differences in measurement methods:

$$PCC_k = \frac{t_k}{\sqrt{t_k^2 + df_k}}$$

with standard error:

$$SE_k = \sqrt{\left(1 - PCC_k^2\right) / df_k}$$

- PCC thresholds are ± 0.104, 0.226, and 0.386 for small, medium, and large effects (Doucouliagos, 2011).

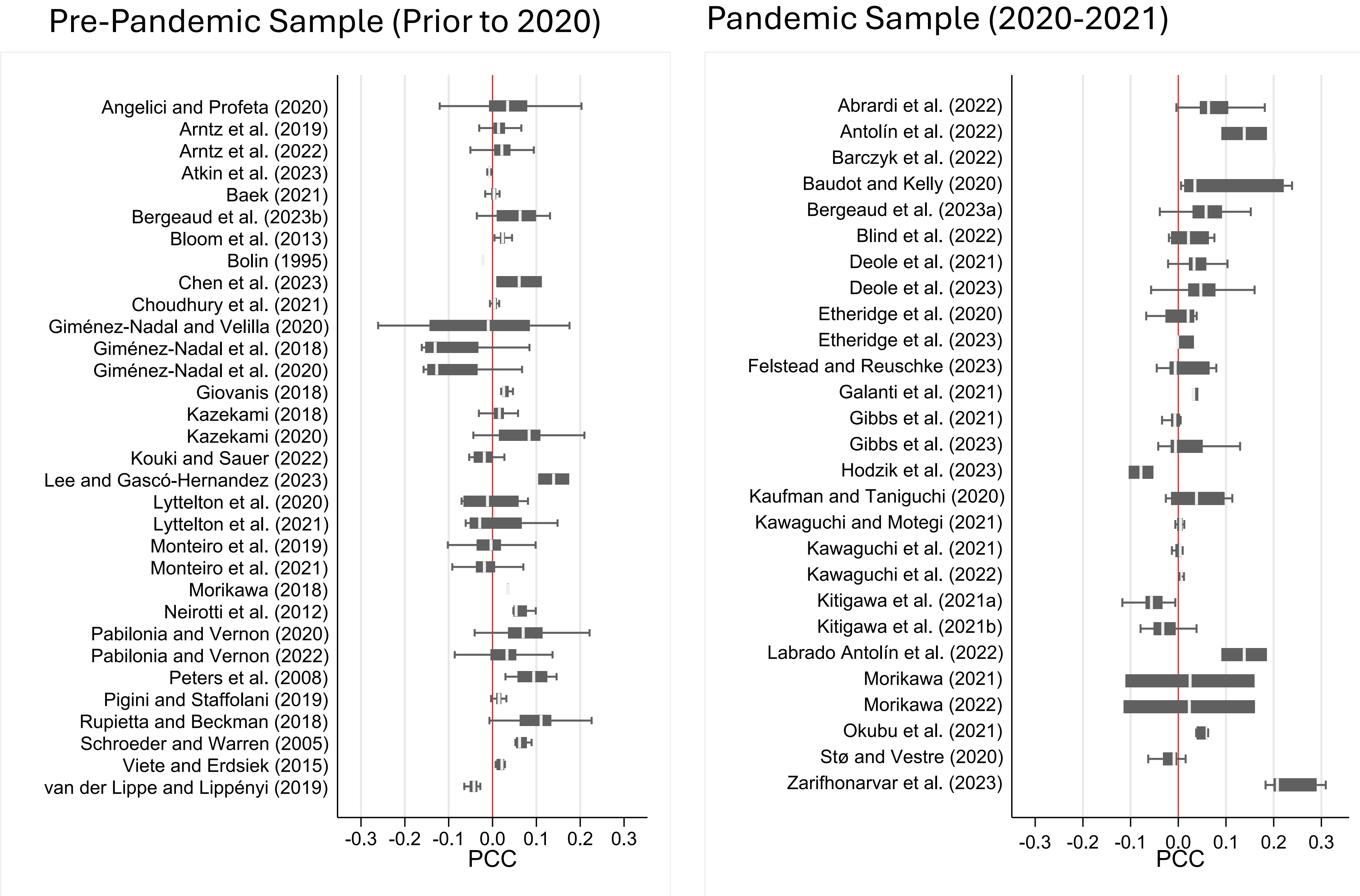


Figure 1. Box-plot of partial correlation coefficients (*PCC*) by study

Table 1. Unconditional meta-averages

	Pre-Pandemic Sample (Prior to 2020)		Pandemic Sample (2020 – 2021)	
	Number of estimates (k)	Random effects [a] (z value)	Number of estimates (k)	Random effects [a] (z value)
All studies	1,740	0.018*** (18.07)	948	0.037*** (20.95)
WFH frequency				
WFH Exclusively	140	-0.001 (-0.46)	221	0.018*** (4.82)
WFH Non-exclusively	935	0.023*** (23.12)	153	0.026*** (6.62)
Sex				
Male	447	0.025*** (12.66)	219	0.039*** (16.22)
Female	459	0.016*** (7.60)	219	0.048*** (17.96)

Notes:
[a] Null hypothesis: The synthesized effect size is zero.
***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Meta-Synthesis

Table 1 (Continued). Unconditional meta-averages

	Pre-Pandemic Sample (Prior to 2020)		Pandemic Sample (2020 – 2021)	
	Number of estimates (k)	Random effects [a] (z value)	Number of estimates (k)	Random effects [a] (z value)
Parental status				
Parents	337	0.034*** (21.93)	92	0.022*** (5.81)
Non-parents	270	0.007*** (7.18)	55	0.052*** (8.19)
Sector				
ICT	45	-0.001 (-0.28)	99	0.012* (1.92)
Manufacturing	11	0.008 (0.81)	38	-0.040*** (-6.93)
Public sector	85	0.004*** (3.04)		
Services (excluding ICT)	78	0.023*** (8.51)	25	-0.015*** (-2.78)

Notes:

[a] Null hypothesis: The synthesized effect size is zero.

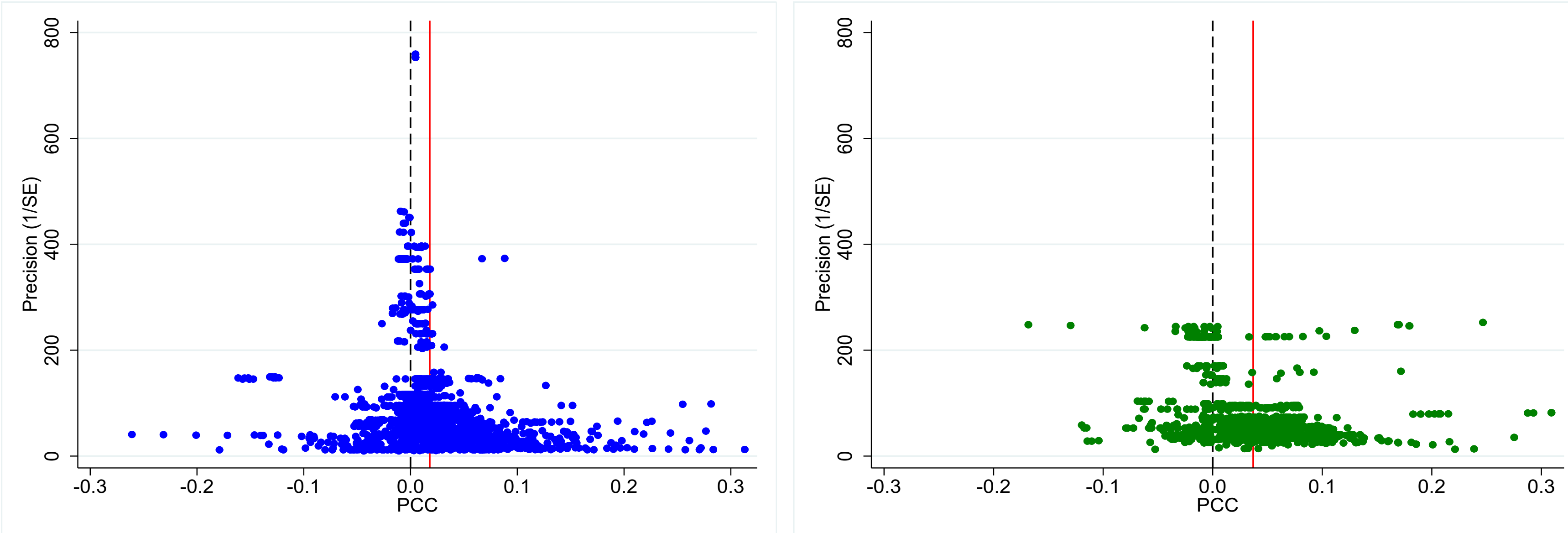
***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Publication Selection Bias

- Publication selection bias (*PSB*) occurs when editors/reviewers reject, or scholars intentionally suppress results that:
  a) Are not statistically significant
  b) Are opposite in sign to the expected result or conventional theory

- A failure to observe suppressed results would mean the observed estimates of Equation (1) do not provide a representative draw from the true sampling distribution of $PCCs$; analysis based only on observed literature would therefore produce a biased estimate of the true parameter.

Figure 2: Funnel Plots



Pre-Covid Era Studies (Prior to 2020) | Covid Era Studies (2020-Present)

- We perform the FAT and PET-PEESE (Stanley and Doucouliagos, 2014) using OLS, WLS, REML, and IV.
- We also employ the Selection Model (Andrews and Kasy, 2019), p-uniform* (van Aert and van Assen, 2018), the Kink model (Bom and Rachinger, 2019), and the Stem method (Furukawa, 2019).

PSB Analysis

Meta-Regression Analysis

- We estimate:

$$PCC_k = \beta_0 + \gamma SE_k^2 + \beta_1 Z_k + \nu_k$$

  where:
  - The inclusion of SE_k^2 both measures the extent of and adjusts for the presence of PSB
  - Z controls for differences in sample characteristics and study methodologies
  - ν is an error term
- Because of uncertainty over which variables to include in Z, we use Bayesian Model Averaging (BMA).

Bayesian Model Averaging

- Estimates multiple candidate models using different combinations of potential explanatory variables in Z.

- Based on model fit, each model is assigned a Posterior Model Probability (**PMP**) indicating the likelihood that it represents the true model specification.

- The PMPs are used as weights to construct a weighted average – the **posterior mean** – of the estimated coefficients for each variable. The PMPs from each model where the variable appears are summed to produce the posterior inclusion probability (**PIP**). A **PIP > 50%** indicates that the variable likely belongs in the true model specification.

BMA Explanation

WFH and Employees' Productivity

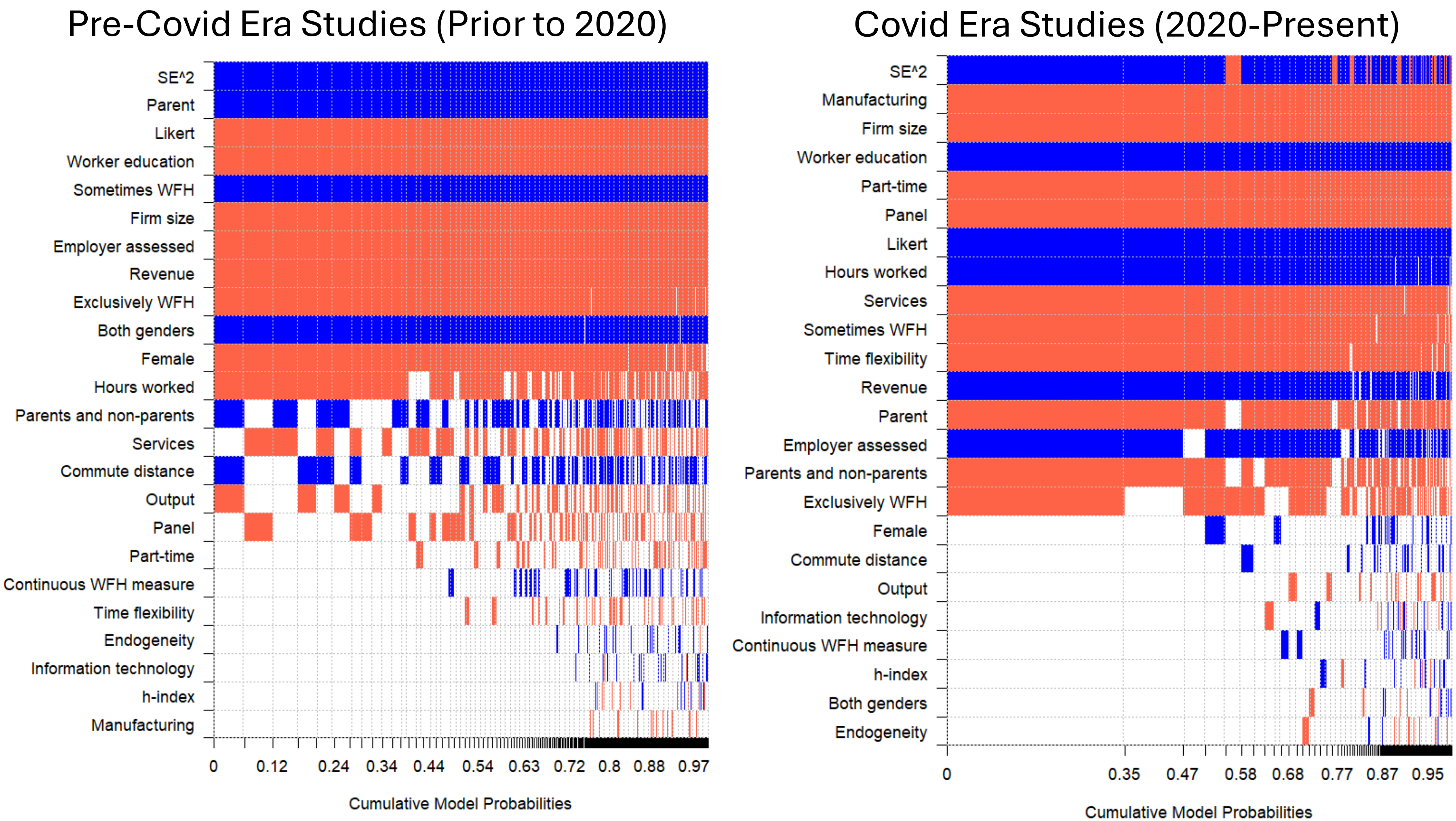


Figure 3: Model Inclusion in Bayesian Model Averaging

Table 2. Analysis of working from home and productivity

Dependent variable: PCC	Pre-Pandemic Sample					Pandemic Sample				
	(a) BMA			(b) OLS (Frequentist check)		(c) BMA			(d) OLS (Frequentist check)	
Meta-independent variable	PIP	Mean	SD	Coeff.	p-value	PIP	Mean	SD	Coeff.	p-value
Publication selection bias										
SE^2	--[a]	5.163	1.240	6.007	0.026	--[a]	0.520	2.870	0.409	0.953
WFH frequency (reference: WFH any amount)										
WFH Non-exclusively	1.000	0.023	0.004	0.023	0.043	0.989	-0.018	0.005	-0.018	0.000
WFH Exclusively	0.996	-0.033	0.008	-0.034	0.005	0.717	-0.009	0.006	-0.012	0.087
Sex (reference: studies of males)										
Studies of females	0.985	-0.012	0.003	-0.012	0.003	0.106	0.001	0.002		
Studies combining males and females	0.994	0.021	0.005	0.019	0.126	0.035	0.000	0.001		
Parental Status (reference: studies of non-parents)										
Studies of parents	1.000	0.022	0.004	0.024	0.000	0.922	-0.028	0.012	-0.031	0.002
Studies combining parents and non-parents	0.538	0.006	0.007	0.012	0.099	0.866	-0.018	0.009	-0.021	0.001
Sector (refence: studies of multiple sectors)										
Studies of the ICT sector	0.034	0.000	0.003			0.062	0.000	0.003		
Studies of the manufacturing sector	0.026	0.000	0.003			1.000	-0.110	0.009	-0.109	0.000
Studies of the public sector	0.056	0.001	0.004							
Studies of the services sector (excluding ICT)	0.490	-0.009	0.010			0.994	-0.064	0.017	-0.065	0.001
Number of observations	1,740					948				
Number of studies	32					27				

[a] *SE^2* is included in all models to adjust for the possibility of PSB. Treating *SE^2* as a candidate variable produces of PIP of 0.991 for the pre-pandemic sample and 0.032 for the pandemic sample. Results indicate that the pre-pandemic sample suffers from PSB, but the pandemic sample likely does not.

Table 2 (Continued). Analysis of working from home and productivity

Dependent variable: PCC	Pre-Pandemic Sample					Pandemic Sample				
	(a) BMA			(b) OLS (Frequentist check)		(c) BMA			(d) OLS (Frequentist check)	
Meta-independent variable	PIP	Mean	SD	Coeff.	p-value	PIP	Mean	SD	Coeff.	p-value
Study characteristics										
Productivity measure (reference: income earned)										
Likert scale	1.000	-0.075	0.009	-0.070	0.018	1.000	0.053	0.007	0.053	0.000
Output produced	0.367	-0.009	0.014			0.067	-0.001	0.006		
Revenue	1.000	-0.046	0.011	-0.041	0.004	0.970	0.053	0.017	0.053	0.017
Total hours worked	0.818	-0.009	0.005	-0.009	0.446	0.996	0.048	0.007	0.048	0.024
WFH measure (reference: binary measure of participation in WFH)										
Continuous measure of time spent in WFH	0.126	0.002	0.005			0.056	0.000	0.002		
Endogeneity (reference: WFH participation assumed exogenous)										
Endogenous	0.049	0.000	0.001			0.034	0.000	0.002		
Selection of controls										
Commute distance	0.462	0.005	0.006			0.068	0.000	0.002		
Firm size	1.000	-0.025	0.005	-0.023	0.078	1.000	-0.107	0.009	-0.109	0.000
Worker education	1.000	-0.032	0.005	-0.032	0.063	1.000	0.086	0.009	0.087	0.001
Time flexibility	0.093	-0.001	0.005			0.988	-0.023	0.006	-0.024	0.001
Part-time employee	0.127	-0.001	0.003			1.000	-0.092	0.011	-0.092	0.001
Assessment of productivity (reference: employee self-assessed)										
Employer assessed	1.000	-0.037	0.007	-0.044	0.042	0.894	0.026	0.012	0.027	0.123
Data type (reference: cross-section data)										
Panel data	0.365	-0.003	0.005			1.000	-0.023	0.004	-0.023	0.026
ln(h-index)	0.027	0.000	0.000			0.041	0.000	0.000		
Intercept	1.000	0.052	--	0.047	0.070	1.000	0.049	--	0.053	0.002
Number of observations	1,740					948				
Number of studies	32					27				

Table 3. Conditional point estimates

	Pre-Pandemic Sample		Pandemic Sample	
	Estimate	[90% Interval]	Estimate	[90% Interval]
All studies	0.022	[0.017; 0.027]	0.039	[0.036; 0.042]
WFH frequency				
WFH Non-exclusively	0.035	[0.023; 0.047]	0.026	[0.019; 0.032]
WFH Exclusively	-0.022	[-0.040; -0.004]	0.032	[0.025; 0.040]
Sex				
Females	0.004	[-0.005; 0.012]		
Males	0.016	[0.007; 0.025]		
Parental status				
Parents	0.034	[0.026; 0.041]	0.028	[0.015; 0.040]
Non-parents	0.009	[0.002; 0.017]	0.059	[0.052; 0.067]
Sector				
Manufacturing			-0.065	[-0.090; -0.039]
Services (Excluding ICT)			-0.020	[-0.048; 0.009]
Other (Including Multi-sector estimates)			0.045	[0.041; 0.049]

Notes: Point estimates obtained from the fitted values of the OLS estimation reported in Table 2 employing the variables selected by Bayesian model averaging as having a PIP > 0.5. Fitted values are calculated using the averages of all the included variables with the exception of the variable of interest, which is set to 1, and any other variables within the same variable group, which are set to Zero. The 90% confidence interval is calculated using standard errors clustered by study.

Conclusions

1. WFH increases employee productivity but:
   - Productivity gains are small.
   - Productivity changes are not uniform across workers.

2. Prior to the pandemic, males benefited more than females, but there was no difference during the pandemic. Parents benefitted more pre-pandemic, but non-parents benefitted more during the pandemic.

3. No difference across sectors before the pandemic, but manufacturing saw productivity losses from WFH during the pandemic.

Recommendations

- Under the "normal" pre-pandemic circumstances, employees who worked from home non-exclusively outperformed those who worked from home exclusively – no studies post-pandemic so someone get on that…
- Measurement matters – future researchers should use more than measure of productivity to demonstrate the robustness of their results.
- Future studies should control for firm size, worker education, whether the work has time flexibility to adjust hours of work, and whether the worker is a part-time employee to avoid omitted variable bias.
- Future studies should use employer reported data since employees may understate or overstate their actual output level.

Meta-Synthesis

$$PCC_k = \beta_0 + \nu_k$$

Meta Fixed-Effects Model:

$$w_k = \frac{1}{SE_k^2}$$

Cochrane's Q test and the I^2 and the H statistics to identify the presence of study-level heterogeneity

Meta Random-Effects Model:

$$w_k = \frac{1}{SE_k^2 + \tau^2}$$

where τ measures the between-study heterogeneity variance (see Borenstein et al., 2021).

Meta Averages

WFH and Employees' Productivity

Meta-Synthesis

Meta Averages

	Number of estimates (k)	Fixed effects [a] (z value)	Random effects [a] (z value)	Cochran Q test of homogeneity [b]	I^2 statistic [c]	H statistic [d]
(a) Pre-COVID era WFH studies						
All studies	1,740	0.006***	0.018***	19,520.69***	90.87	10.95
		(34.43)	(18.07)	(0.000)		
WFH frequency						
WFH Exclusively	140	0.001**	-0.001	3,148.46***	94.91	17.20
		(2.54)	(-0.46)	(0.000)		
WFH Non-exclusively	935	0.015***	0.023***	4,080.05***	77.11	4.37
		(43.72)	(23.12)	(0.000)		
Gender						
Male	447	0.015***	0.025***	5,246.97***	91.50	11.76
		(26.99)	(12.66)	(0.000)		
Female	459	0.006***	0.016***	5,725.89***	92.00	12.50
		(10.26)	(7.60)	(0.000)		
Parental status						
Parents	337	0.032***	0.034***	1,092.36***	69.24	3.25
		(39.11)	(21.93)	(0.000)		
Non-parents	270	0.006***	0.007***	571.39***	52.92	2.12
		(8.55)	(7.18)	(0.000)		
Sector						
Information and communications technology	45	-0.002***	-0.001	2,006.49***	97.81	45.60
		(-5.65)	(-0.28)	(0.000)		
Manufacturing	11	0.005	0.008	26.06***	61.63	2.61
		(0.83)	(0.81)	(0.004)		
Public sector	85	0.003***	0.004***	724.10***	82.32	5.66
		(11.28)	(3.04)	(0.000)		
Services (excluding ICT)	78	0.022***	0.023***	351.35***	78.08	4.56
		(21.56)	(8.51)	(0.000)		

WFH and Employees' Productivity

Meta-Synthesis

Meta Averages

	Number of estimates (k)	Fixed effects [a] (z value)	Random effects [a] (z value)	Cochran Q test of homogeneity [b]	I^2 statistic [c]	H statistic [d]
(b) COVID era WFH studies						
All studies	948	0.020***	0.037***	28,752.51***	96.70	30.27
		(56.58)	(20.95)	(0.000)		
WFH frequency						
WFH Exclusively	221	0.013***	0.018***	20,214.28***	98.91	91.88
		(29.35)	(4.82)	(0.000)		
WFH Non-exclusively	153	0.019***	0.026***	1,096.69***	86.14	7.22
		(12.66)	(6.62)	(0.000)		
Gender						
Male	219	0.020***	0.039***	898.25***	75.73	4.12
		(19.41)	(16.22)	(0.000)		
Female	219	0.027***	0.048***	1,734.01***	87.43	7.95
		(28.00)	(17.96)	(0.000)		
Parental status						
Parents	92	0.009***	0.022***	1,028.46***	91.15	11.30
		(8.36)	(5.81)	(0.000)		
Non-parents	55	0.017***	0.052***	2,002.73***	97.30	37.09
		(16.47)	(8.19)	(0.000)		
Sector						
Information and communications technology	99	0.013***	0.012*	19,414.34***	99.50	198.11
		(28.06)	(1.92)	(0.000)		
Manufacturing	38	-0.044***	-0.040***	164.72***	77.54	4.45
		(-18.06)	(-6.93)	(0.000)		
Services (excluding ICT)	25	-0.016***	-0.015***	107.87***	77.75	4.49
		(-6.33)	(-2.78)	(0.000)		

PSB Analysis

Funnel Plots

	Pre-COVID era studies				COVID era studies			
Panel (a) FAT-PET analysis	**OLS**	**WLS**	**REML**	**IV**	**OLS**	**WLS**	**REML**	**IV**
Presence of PB (FAT)	0.534***	1.197***	0.144	0.547***	1.227***	1.873***	0.581***	1.260***
Null hypothesis: PB is not present	(0.084)	(0.298)	(0.108)	(0.150)	(0.241)	(0.351)	(0.202)	(0.296)
Estimate of the nexus (PET)	0.010***	-0.000	0.017*	0.010**	0.014***	0.004	0.025*	0.014
Null hypothesis: nexus is zero	(0.002)	(0.002)	(0.009)	(0.004)	(0.004)	(0.004)	(0.013)	(0.009)
Number of observations	1,740	1,740	1,740	1,740	948	948	948	948
Number of studies	32	32	32	32	27	27	27	27
Panel (b) PEESE analysis	**OLS**	**WLS**	**REML**	**IV**	**OLS**	**WLS**	**REML**	**IV**
Presence of PB	5.012***	19.325**	-0.837	5.242***	18.654***	49.354***	13.533***	19.308***
Null hypothesis: PB is not present	(1.044)	(8.293)	(1.243)	(1.607)	(4.869)	(7.679)	(3.560)	(6.711)
Estimate of the nexus	0.017***	0.005**	0.021**	0.017***	0.029***	0.014***	0.026**	0.029***
Null hypothesis: nexus is zero	(0.001)	(0.003)	(0.009)	(0.004)	(0.003)	(0.004)	(0.013)	(0.008)
Number of observations	1,740	1,740	1,740	1,740	948	948	948	948
Number of studies	32	32	32	32	27	27	27	27
Panel (c) Additional non-linear techniques	**Selection Model**	**p-uniform***	**Kink model**	**Stem model**	**Selection Model**	**p-uniform***	**Kink model**	**Stem model**
Estimate of the nexus	0.023***	0.029***	-0.0003	0.005	0.042***	0.030***	0.004	0.195*
Null hypothesis: nexus is zero	(0.002)	(0.001)	(0.001)	(0.004)	(0.003)	(0.003)	(0.003)	(0.107)
Number of observations	1,740	1,740	1,740	1,740	948	948	948	948
Number of studies	32	32	32	32	27	27	27	27

Bayesian Model Averaging

- For J candidate variables:

$$PCC_k = \beta_0 + \sum_{j=1}^{J} \beta_j Z_{j,k} + \nu_k$$

- Starting from the Bayes rule, the posterior probability density is given by the following:

$$p(\beta|PCC,Z) = \frac{p(PCC,Z|\beta)p(\beta)}{p(PCC,Z)}$$

where $p(PCC,Z|\beta)$ is the marginal likelihood, $p(\beta)$ is the prior probability density, and $p(PCC,Z)$ is the probability of the data.

- With J potential regressors, there are 2^J potential models M. There are therefore $M_1, \dots, M_\mu$ models where $\mu \in [1, 2^J]$.

- Assuming a likelihood function and a prior probability density, the posterior probability density for each model M_μ is written as:

$$p(\beta_\mu|M_\mu,PCC,Z) = \frac{p(PCC|\beta_\mu,M_\mu,Z)p\left(\beta_\mu\middle|M_\mu\right)}{p(PCC|M_\mu,Z)}$$

$$PMP = p(M_\mu|PCC,Z) = \frac{p(PCC|M_\mu,Z)p(M_\mu)}{\sum_{\mu=1}^{2^J} p(Z|M_\mu)\, p(M_\mu)}$$

$$PIP_j = \sum_{\mu=1}^{2^J} p(M_\mu|PCC)$$

- We use the unit information g prior for the parameters and the uniform prior for the model space.

- As robustness checks, we use the BRIC prior and hyper-g prior for parameters and the beta-binomial prior for the model space.

Methodology