Pandemic programming

How COVID-19 affects software developers and how their organizations can help

Paul Ralph[1] ⬤ · Sebastian Baltes[2] · Gianisa Adisaputri[1] · Richard Torkar[3,4] ·
Vladimir Kovalenko[5] · Marcos Kalinowski[6] · Nicole Novielli[7] · Shin Yoo[8] ·
Xavier Devroey[9] · Xin Tan[10] · Minghui Zhou[10] · Burak Turhan[11,12] · Rashina Hoda[11] ·
Hideaki Hata[13] · Gregorio Robles[14] · Amin Milani Fard[15] · Rana Alkadhi[16]

Abstract

Context As a novel coronavirus swept the world in early 2020, thousands of software developers began working from home. Many did so on short notice, under difficult and stressful conditions.

Objective This study investigates the effects of the pandemic on developers' wellbeing and productivity.

Method A questionnaire survey was created mainly from existing, validated scales and translated into 12 languages. The data was analyzed using non-parametric inferential statistics and structural equation modeling.

Results The questionnaire received 2225 usable responses from 53 countries. Factor analysis supported the validity of the scales and the structural model achieved a good fit (CFI = 0.961, RMSEA = 0.051, SRMR = 0.067). Confirmatory results include: (1) the pandemic has had a negative effect on developers' wellbeing and productivity; (2) productivity and wellbeing are closely related; (3) disaster preparedness, fear related to the pandemic and home office ergonomics all affect wellbeing or productivity. Exploratory analysis suggests that: (1) women, parents and people with disabilities may be disproportionately affected; (2) different people need different kinds of support.

✉ Paul Ralph
  paulralph@dal.ca

✉ Sebastian Baltes
  sebastian.baltes@adelaide.edu.au

✉ Burak Turhan
  burak.turhan@monash.edu

Extended author information available on the last page of the article.

Conclusions To improve employee productivity, software companies should focus on maximizing employee wellbeing and improving the ergonomics of employees' home offices. Women, parents and disabled persons may require extra support.

1 Introduction

In December 2019, a novel coronavirus disease (COVID-19) emerged in Wuhan, China. While symptoms vary, COVID-19 often produces fever, cough, shortness of breath, and in some cases, pneumonia and death. By April 30, 2020, The World Health Organization (WHO) recorded more than 3 million confirmed cases and 217,769 deaths (WHO 2020a). With wide-spread transmissions in 214 countries, territories or areas, the WHO declared it a *Public Health Emergency of International Concern* (WHO 2020b) and many jurisdictions declared states of emergency or lockdowns (Kaplan et al. 2020). Many technology companies told their employees to work from home (Duffy 2020).

Thinking of this situation as a global natural experiment in working from home—the event that would irrefutably verify the benefits of working from home—would be naïve. This is not normal working from home. This is attempting to work from home, unexpectedly, during an unprecedented crisis. The normal benefits of working from home may not apply (Donnelly and Proctor-Thomson 2015). Rather than working in a remote office or well-appointed home office, some people are working in bedrooms, at kitchen tables and on sofas while partners, children, siblings, parents, roommates, and pets distract them. Others are isolated in a studio or one-bedroom apartment. With schools and childcare closed, many parents juggle work with not only childcare but also home schooling or monitoring remote schooling activities and keeping children engaged. Some professionals have the virus or are caring for ill family members.

> *Quarantine work !== Remote work. I've been working remotely with success for 13 years, and I've never been close to burn out. I've been working quarantined for over a month and I'm feeling a tinge if burn out for the first time in my life. Take care of yourself folks. Really.*
>
> –Scott Hanselman (@shanselman), April 20, 2020

While numerous studies have investigated remote work, few investigate working from home during disasters. There are no modern studies of working from home during a pandemic of this magnitude because there has not been a pandemic of this magnitude since before there was a world wide web. Therefore, software companies have limited evidence on how to support their workers through this crisis, which raises the following research question.

Research question: How is working from home during the COVID-19 pandemic affecting software developers' emotional wellbeing and productivity?

To address this question, we generate and evaluate a theoretical model for explaining and predicting changes in wellbeing and productivity while working from home during

a crisis. Moreover, we provide recommendations for professionals and organizations to support employees who are working from home due to COVID-19 or future disasters.

2 Background

To fully understand this study, we need to review several areas of related work: pandemics, bioevents and disasters; working from home; and productivity and wellbeing.

2.1 Pandemics, Bioevents and Disasters

Madhav et al. (2017) defines pandemics as "large-scale outbreaks of infectious disease over a wide geographic area that can greatly increase morbidity and mortality and cause significant economic, social, and political disruption" (p. 35). Pandemics can be very stressful not only for those who become infected but also for those caring for the infected and worrying about the health of themselves, their families and their friends (Kim et al. 2015; Prati et al. 2011). In a recent poll, "half of Canadians (50%) report[ed] a worsening of their mental health" during the COVID-19 lockdown (ARI 2020). In Australia, the pandemic appears to have doubled the incidence of mental health problems (Fisher et al. 2020).

A pandemic can be mitigated in several ways including *social distancing* (Anderson et al. 2020): "a set of practices that aim to reduce disease transmission through physical separation of individuals in community settings" (p. 120-14 Rebmann 2009), including public facility shutdowns, home quarantine, cancelling large public gatherings, working from home, reducing the number of workers in the same place at the same time and maintaining a distance of at least 1.5–2m between people (Rebmann 2009; Anderson et al. 2020).

The extent to which individuals comply with recommendations varies significantly and is affected by many factors. People are more likely to comply when they have more self-efficacy; that is, confidence that they can stay at home or keep working during the pandemic, and when they perceive the risks as high (Teasdale et al. 2012). However, this "threat appraisal" depends on: the psychological process of quantifying risk, sociocultural perspectives (e.g. one's worldview and beliefs; how worried one's friends are), "illusiveness of preparedness" (e.g. fatalistic attitudes and denial), beliefs about who is responsible for mitigating risks (e.g. individuals or governments) and how prepared one feels (Yong et al. 2017, 2019; Prati et al. 2011).

People are less likely to comply when they are facing loss of income, personal logistical problems (e.g. how to get groceries), isolation, and psychological stress (e.g. fear, boredom, frustration, stigma) (DiGiovanni et al. 2004). Barriers to following recommendations include job insecurity, lack of childcare, guilt and anxiety about work not being completed, and the personal cost of following government advice (Teasdale et al. 2012; Blake et al. 2010).

For employees, experiencing negative life events such as disasters is associated with absenteeism and lower quality of workdays (North et al. 2010). Employers therefore need work-specific strategies and support for their employees. Employers can give employees a sense of security and help them return to work by continuing to pay full salaries on time, reassuring employees they they are not going to lose their job, having flexible work demands, implementing an organized communication strategy, and ensuring access to utilities (e.g. telephone, internet, water, electricity, sanitation) and organisational resources (North et al. 2010; Donnelly and Proctor-Thomson 2015; Blake et al. 2010).

Work-specific strategies and support are also needed to ensure business continuation and survival. The disruption of activities in disasters simultaneously curtails revenues and

reduces productive capacity due to the ambiguity and priorities shifting in individuals, organizations and communities (Donnelly and Proctor-Thomson 2015). As social distancing closes worksites and reduces commerce, governments face increased economic pressure to end social distancing requirements prematurely (Loose et al. 2010). Maintaining remote workers' health and productivity is therefore important for maintaining social distancing as long as is necessary (Blake et al. 2010).

As we prepare this article, many other studies of the COVID-19 pandemic's effects are underway. Early evidence suggests complicated effects on productivity, which vary by person, project and metric (Bao et al. 2020). Some evidence suggests programmers are working longer hours, at an unsustainable pace (Forsgren 2020).

2.2 Working from Home

Pérez et al. (2004) defines *teleworking* (also called *remote working*) as "organisation of work by using information and communication technologies that enable employees and managers to access their labour activities from remote locations" (p. 280). It includes working from home, a satellite office, a telework centre or even a coffee shop. Remote working can help restore and maintain operational capacity and essential services during and after disasters (Blake et al. 2010), especially when workplaces are inaccessible. Indeed, many executives are already planning to shift "at least 5% of previously on-site employees to permanently remote positions post-COVID 19" (Lavelle 2020).

However, many organizations lack appropriate plans, supportive policies, resources or management practices for practising home-based telework. In disasters such as pandemics where public facilities are closed and people are required to stay at home, their experience and capacity to work can be limited by lack of dedicated workspace at home, caring responsibilities and organizational resources (Donnelly and Proctor-Thomson 2015).

In general, working from home is often claimed to improve productivity (Davenport and Pearlson 1998; McInerney 1999; Cascio 2000) and teleworkers consistently report increased perceived productivity (Duxbury et al. 1998; Baruch 2000). Interestingly, Baker et al. (2007) found that organizational and job-related factors (e.g. management culture, human resources support, structure of feedback) are more likely to affect teleworking employees' satisfaction and perceived productivity than work styles (e.g. planning vs. improvising) and household characteristics (e.g. number of children). While increasing productivity, "working from home is associated with greater levels of both work pressure and work-life conflict" (Russell et al. 2009) because work intrudes into developers' home lives through working unpaid overtime, thinking about work in off hours, exhaustion and sleeplessness (Hyman et al. 2003).

Moreover, individuals' wellbeing while working remotely is influenced by their emotional stability (that is, a person's ability to control their emotions when stressed). For people with high emotional stability, working from home provides more autonomy and fosters wellbeing; however, for employees with low emotional stability, it can exacerbate physical, social and psychological strain (Perry et al. 2018). The COVID-19 pandemic has not been good for emotional stability (ARI 2020).

Research on working from home has been criticized for relying on self-reports of perceived productivity, which may inflate its benefits (Bailey and Kurland 2002); however, objective measures often lack construct validity (Ralph and Tempero 2018) and perceived productivity correlates well with managers' appraisals (Baruch 1996). (The perceived productivity scale we use below correlates well with objective performance data; cf. Section 4.2).

2.3 Productivity and Wellbeing

Previous studies suggest that productivity affects project outcomes and is affected by numerous factors including team size and technologies used (McLeod and MacDonell 2011). However, existing research on developer productivity is rife with construct validity problems.

Productivity is the amount of work done per unit of time. Measuring time is simple but quantifying the work done by a software developer is not. Some researchers (e.g. Jaspan and Sadowski 2019) argue for using goal-specific metrics. Others reject the whole idea of measuring productivity (e.g. Ko AJ 2019) not least because people tend to optimize for whatever metric is being used—a phenomenon known as *Goodhart's law* (Goodhart 1984; Chrystal and Mizen 2003).

Furthermore, simple productivity measures such as counting commits or modified lines of code in a certain period suffer from low construct validity (Ralph and Tempero 2018). The importance and difficulty of a commit does not necessarily correlate with its size. Similarly, some developers might prefer dense, one-line solutions while others like to arrange their contributions in several lines.

Nevertheless, large companies including Microsoft still use controversial metrics such as number of pull requests as a "proxy for productivity" (Spataro 2020), and individual developers also use them to monitor their own performance (Baltes and Diehl 2018). Copious time tracking tools exist for that purpose—some specifically tailored for software developers.[1]

While researchers have adapted existing scales to measure related phenomena like happiness (e.g. Graziotin and Fagerholm 2019), there is no widespread consensus about how to measure developers' productivity or the main antecedents thereof. Many researchers use simple, unvalidated productivity scales; for example, Meyer et al. (2017) used a single question asking participants to rate themselves from "not productive" to "very productive." (The perceived productivity scale we use below has been repeatedly validated in multiple domains; cf. Section 4.2).

Meanwhile, a programmer's productivity is closely related to their job satisfaction (Storey et al. 2019) and emotional state (Wrobel 2013; Graziotin et al. 2015). Unhappiness, specifically, leads to "low cognitive performance, mental unease or disorder, and low motivation" (Graziotin et al. 2017, p. 44). However, little is known about the antecedents or consequences of software professionals' physical or mental wellbeing in general.

3 Hypotheses

The related work discussed above suggests numerous hypotheses. Here we hypothesize about "developers" even though our survey was open to all software professionals because most respondents were developers (see Section 5.3). These hypotheses were generated contemporaneously with questionnaire design—before data collection began.

Hypothesis H1: Developers will have lower wellbeing while working from home due to COVID-19. Stress, isolation, travel restrictions, business closures and the absence of educational, child care and fitness facilities all take a toll on those working from home. Indeed, a pandemic's severity and the uncertainty and isolation it induces create frustration, anxiety and fear (Taha et al. 2014; DiGiovanni et al. 2004; Teasdale et al. 2012). It therefore seems likely that many developers will experience reduced emotional wellbeing.

---

[1]e.g. https://wakatime.com/

Hypothesis H2: Developers will have lower perceived productivity while working from home due to COVID-19. Similarly, stress, moving to an impromptu home office, and lack of child care and other amenities may reduce many developers' productivity. Many people are likely more distracted by the people they live with and their own worrisome thoughts. People tend to experience lower motivation, productivity and commitment while working from home in a disaster situation (Donnelly and Proctor-Thomson 2015).

Assuming Hypotheses H1 and H2 are supported, we want to propose a model that explains and predicts changes in wellbeing and productivity (Fig. 1). Hypotheses H1 and H2 are encapsulated in the *change in wellbeing* and *change in perceived productivity* constructs. The model only makes sense if wellbeing and productivity have changed since developers began working from home.

Hypothesis H3: Change in wellbeing and change in perceived productivity are directly related. We expect wellbeing and productivity to exhibit reciprocal causality. That is, as we feel worse, we become less productive, and feeling less productive makes us feel even worse, in a downward spiral. Many studies show that productivity and wellbeing covary (cf. Dall'Ora et al. 2016). Moreover, Evers et al. (2014) found that people with increasing health risks have lower wellbeing and life satisfaction, leading to higher rates of depression and anxiety. Conversely, decreasing health risk will increase physical and emotional wellbeing and productivity.

Hypotheses H4 and H5: Disaster preparedness is directly related to change in wellbeing and change in perceived productivity. Disaster preparedness is the degree to which a person is ready for a natural disaster. It includes behaviors like having an emergency supply kit and complying with directions from authorities. We expect lack of preparedness for disasters in general and for COVID-19 in particular to exacerbate reductions in wellbeing and productivity, and vice versa (cf. Paton 2008; Donnelly and Proctor-Thomson 2015).

Hypotheses H6 and H7: Fear (of the pandemic) is inversely related to change in wellbeing and change in perceived productivity. Fear is a common reaction to bioevents like pandemics. Emerging research on COVID-19 is already showing a negative effect on wellbeing, particularly anxiety (Harper et al. 2020; Xiang et al. 2020). Meanwhile, fear of infection and public health measures cause psychosocial distress, increased absenteeism and reduced productivity (Shultz et al. 2016; Thommes et al. 2016).

Hypotheses H8 and H9: Home office ergonomics is directly related to change in wellbeing and change in perceived productivity. Here we use ergonomics in its broadest sense of the degree to which an environment is safe, comfortable and conducive to the tasks being completed in it. We are not interested in measuring the angle of a developer's
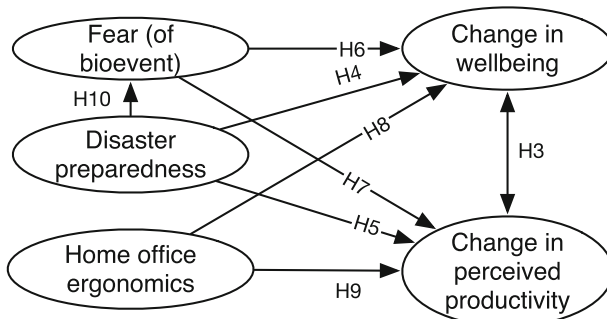

Fig. 1 Theoretical model of developer wellbeing and productivity

knees and elbows, but in a more general sense of their comfort. Professionals with more ergonomic home offices should have greater wellbeing and be more productive. Donnelly and Proctor-Thomson (2015) found that availability of a dedicated work-space at home, living circumstances, and the availability of organisational resources to work relate to the capacity to return to work after a disaster and employees' productivity.

Hypothesis H10: Disaster preparedness is inversely related to fear (of the pandemic). It seems intuitive that the more prepared we are for a disaster, the more resilient and less afraid we will be when the disaster occurs. Indeed, Ronan et al.'s 2015 systematic review found that programs for increasing disaster preparedness had a small- to medium-sized negative effect on fear. People who have high self-efficacy and response-efficacy (i.e. perceive themselves as ready to face a disaster) will be less afraid (Roberto et al. 2009).

4 Method

On March 19, 2020, the first author initiated a survey to investigate how COVID-19 affects developers, and recruited the second and third authors for help. We created the questionnaire and it was approved by Dalhousie University's research ethics board in less than 24 hours. We began data collection on March 27th. We then recruited authors 5 through 17, who translated and localized the questionnaire into Arabic, (Mandarin) Chinese, English, French, Italian, Japanese, Korean, Persian, Portuguese, Spanish, Russian and Turkish, and created region-specific advertising strategies. Translations launched between April 5 and 7, and we completed data collection between April 12 and 16. Next, we recruited the fourth author to assist with the data analysis, which was completed on April 29. The manuscript was prepared primarily by the first four authors with edits from the rest of team.

This section details our approach and instrumentation.

4.1 Population and Inclusion Criteria

This study's target population is software developers anywhere in the world *who switched from working in an office to working from home* because of COVID-19. Of course, developers who had been working remotely before the pandemic and developers who continued working in offices throughout the pandemic are also important, but this study is about the switch, and the questions are designed for people who switched from working on-site to at home.

In principle, the questionnaire was open to all sorts of software professionals, including designers, quality assurance specialists, product managers, architects and business analysts, but we are mainly interested in developers, our marketing focused on software developers, and therefore most respondents identify as developers (see Section 5.3).

4.2 Instrument Design

We created an anonymous questionnaire survey. We did not use URL tracking or tokens. We did not collect contact information.

Questions were organized into blocks corresponding to scale or question type. The order of the items in each multi-item scale was randomized to mitigate primacy and recency effects. The order of blocks was not randomized because our pilot study (Section 4.3) suggested that the questionnaire was more clear when the questions that distinguish between before and after the switch to home working came after those that do not.

The questionnaire used a filter question to exclude respondents who do not meet the inclusion criteria. Respondents who had not switched from working in an office to working from home because of COVID-19 simply skipped to the end of the questionnaire. It also included not only traditional demographic variables (e.g. age, gender, country, experience, education) but also how many adults and children (under twelve) participants lived with, the extent to which participants are staying home and whether they or any friends or family had tested positive for COVID-19.

The questionnaire used validated scales as much as possible to improve construct validity. A *construct* is a quantity that cannot be measured directly. Fear, disaster preparedness, home office ergonomics, wellbeing and productivity are all constructs. In contrast, age, country, and number of children are all directly measurable. Direct measurements are assumed to have inherent validity, but latent variables have to be validated to ensure that they measure the right properties (cf. Ralph and Tempero 2018).

The exact question wording can be seen in our replication pack (see Section 8). This section describes the scales and additional questions.

Emotional wellbeing (WHO-5) To assess emotional wellbeing, we used the WHO's five-item wellbeing index (WHO-5).[2] Each item is assessed on a six-point scale from "at no time" (0) to "all of the time" (5). The scale can be calculated by summing the items or using factor analysis. The WHO-5 scale is widely used, widely applicable, and has high sensitivity and construct validity (Topp et al. 2015). Respondents self-assessed their wellbeing twice: once for the four weeks prior to beginning to work from home, and then again for the time they have been working from home.

Perceived Productivity (HPQ) To assess perceived productivity we used items from the WHO's Health and Work Performance Questionnaire (HPQ).[3] The HPQ measures perceived productivity in two ways: (1) using an eight-item summative scale, with multiple reversed indicators, that assesses overall and relative performance; and (2) using 11-point general ratings of participants' own performance and typical performance of similar workers. These scales are amenable to factor analysis or summation. Of course, people tend to overestimate their performance relative to their peers, but we are comparing participants to their past selves not to each other. HPQ scores are closely related to objective performance data in diverse fields (Kessler et al. 2003). Again, respondents self-assessed their productivity for both the four weeks prior to working from home, and for the time they have been working from home.

Disaster Preparedness (DP) To assess disaster preparedness, we adapted Yong et al.'s (2017) individual disaster preparedness scale. Yong et al. developed their five-item, five-point, Likert scale based on common, important behaviors such as complying with government recommendations and having emergency supplies. The scale was validated using a questionnaire survey of a "weighted nationally representative sample" of 1084 Canadians. We adapted the items to refer specifically to COVID-19. It can be computed by summing the responses or using factor analysis.

---

[2]https://www.psykiatri-regionh.dk/who-5/who-5-questionnaires/Pages/default.aspx
[3]https://www.hcp.med.harvard.edu/hpq/info.php

Fear and Resilience (FR) The Bracha-Burkle Fear and Resilience (FR) checklist is a triage tool for assessing patients' reactions to bioevents (including pandemics). The FR checklist places the patient on a scale from intense fear to hyper-resilience (Bracha and Burkle 2006). We dropped some of the more extreme items (e.g. "Right now are you experiencing shortness of breath?") because respondents are at home taking a survey, not arriving in a hospital emergency room. The FR checklist is a weighted summative scale so it has to be computed manually using Bracha and Burkle's formula rather than using factor analysis. It has multiple reversed indicators.

Ergonomics We could not find a reasonable scale for evaluating home office ergonomics. There is comparatively less research on the ergonomics of home offices (Inalhan and Ng 2010) and ergonomic instruments tend to be too narrow (e.g. evaluating hip angle). Based on our reading of the ergonomics literature, we made a simple six-item, six-point Likert scale concerning distractions, noise, lighting, temperature, chair comfort and overall ergonomics.

Again, we evaluated the scale's face and content validity using a pilot study (see Section 4.3) and examine convergent and discriminant validity ex post in Section 5.2.

Organizational Support (OS) We could not find any existing instrument that measures the degree to which an organization supports its employees during a crisis. The first author therefore interviewed three developers with experience in both co-located and distributed teams as well as office work and working from home. Interviewees brainstormed actions companies could take to help, and we used open-coding (Saldaña 2015) to organize their responses into five themes:

1. *Equipment:* providing equipment employees need in their home office (e.g. a second monitor)
2. *Reassurance:* adopting a tone that removes doubt and fear (e.g. assuring employees that lower productivity would be understood)
3. *Connectedness:* encouraging virtual socializing (e.g. through video chat)
4. *Self-care:* providing personal services not directly related to work (e.g. resources for exercising or home-schooling children)
5. *Technical infrastructure and practices:* ensuring that remote infrastructure (e.g. VPNs) and practices (e.g. code review) are in place.

We generated a list of 22 actions (four or five per theme) by synthesizing the ideas of interviewees with existing literature on working from home, distributed development and software engineering more generally. For each action, respondents indicate whether their employer is taking the action and whether they think it is or would be helpful. Organizational support is not a construct in our theory per se because we have insufficient a priori information to produce a quantitative estimate, so we analyze these answers separately.

4.3 Pilot

We solicited feedback from twelve colleagues: six software engineering academics and six experienced software developers. Pilot participants made various comments on the questionnaire structure, directions and on the face and content validity of the scales. Based on this feedback we made numerous changes including clarifying directions, making the question order static, moving the WHO-5 and HPQ scales closer to the end, dropping some problematic questions, splitting up an overloaded question, and adding some open response

questions. (Free-text answers are not analyzed in this paper; open response questions were included mainly to inform future research; see Section 6.3).

4.4 Sampling, Localization and Incentives

We advertised our survey on social and conventional media, including Dev.to, Développez.com, DNU.nl, eksisozluk, Facebook, Hacker News, Heise Online, InfoQ, LinkedIn, Twitter, Reddit and WeChat. Upon completion, participants were provided a link and encouraged to share it with colleagues who might also like to take the survey. Because this is an anonymous survey, we did not ask respondents to provide colleagues' email addresses.

We considered several alternatives, including scraping emails from software repositories and stratified random sampling using company lists, but none of these options seemed likely to produce a more representative sample. Granted, if we sampled from an understood sampling frame, we could better evaluate the representativeness of the sample and generalizability of the results; however, we are not aware of any sampling frames with sufficiently well-understood demographics to facilitate accurate inferences.

Instead, we focused on increasing the diversity of the sample by localizing the survey and promoting it in more jurisdictions. We translated the survey into Arabic, (Mandarin) Chinese, French, Italian, Japanese, Korean, Persian, Portuguese, Spanish, Russian and Turkish. Each author-translator translated from English into their first language. We capitalized on each authors' local knowledge to reach more people in their jurisdiction. Rather than a single, global campaign, we used a collection of local campaigns.

Each localization involved small changes in wording. Only a few significant changes were needed. The Chinese version used a different questionnaire system (https://wjx.cn) because Google Forms is not available in China. Furthermore, because the lockdowns in China and Korea were ending, we reworded some questions from "since you began working from home" to "while you were working from home."

We did not offer cash incentives for participation. Rather, we offered to donate US$500 to an open source project chosen by participants (in one of the open response questions). Respondents suggested a wide variety of projects, so we donated US$100 to the five most mentioned: The Linux Foundation, The Wikimedia Foundation, The Mozilla Foundation, The Apache Software Foundation and the Free Software Foundation. The Portuguese version was slightly different: it promised to donate 1000 BRL to Ação da Cidadania's (a Brazillian NGO) Action against Corona project rather than a project chosen by participants (which we did).

5 Analysis and Results

We received 2668 total responses of which 439 did not meet our inclusion criteria and 4 were effectively blank (see below) leaving 2225. This section describes how the data was cleaned and analyzed.

5.1 Data Cleaning

The data was cleaned as follows.

1. Delete responses that do not meet inclusion criteria.

2. Delete almost empty rows, where the respondent apparently answered the filter question correctly, then skipped all other questions.

3. Delete the timestamp field (to preserve anonymity), the consent form confirmation field (because participants could not continue without checking these boxes so the answer is always "TRUE") and the filter question field (because all remaining rows have the same answer).

4. Add a binary field indicating whether the respondent had entered text in at least one long-answer question (see Section 5.2).

5. Move all free-text responses to a separate file (to preserve anonymity).

6. Recode the raw data (which is in different languages with different alphabets) into a common quantitative coding scheme; for example, from 1 for "strongly disagree" to 5 for "strongly agree" The recoding instructions and related scripts are included in our replication package (see Section 8).

7. Split select-multiple questions into one binary variable per checkbox (Google Forms uses a comma-separated list of the text of selected answers).

8. Add a field indicating the language of the response.

9. Combine the responses into a single data file.

10. Calculate the FR scale according to its formula (Bracha and Burkle 2006).

5.2 Validity Analysis

We evaluated construct validity using established guidelines (Ralph and Tempero 2018). First, we assessed content validity using a pilot study (Section 4.3). Next, we assessed convergent and discriminant validity using a principle component analysis (PCA) with Varimax rotation and Kaiser normalization. Bartlett's test is significant ($\chi^2 = 13229; df = 253; p < 0.001$) and our KMO measure of sampling adequacy is high (0.874), indicating that our data is appropriate for factor analysis.

As Table 1 shows, the items load well but not perfectly. The bold coefficients suggest possible issues with Change in Productivity (ΔProductivity) 7 and 9, as well as Ergonomics 1. We dropped items one at a time until the loadings stabilized, starting with ΔProductivity 7, followed by ΔProductivity 9. As shown in Table 2, dropping these two indicators solved the problem with Ergonomics 1, so the latter is retained.

We evaluate predictive validity by testing our hypotheses in Section 5.4.

Response bias Here, *response bias* refers to the possibility that people for whom one of our hypotheses hold are more likely to take the questionnaire, thus inflating the results.

There are two basic ways to analyze this kind of response bias. The first—comparing sample parameters to known population parameters—is impractical because no one has ever established population parameters for software professionals. The second—comparing late respondents to early respondents—cannot be used because we do not know the time between each respondent learning of the survey and completing it. However, we can do something similar: we can compare respondents who answered one or more open response questions (more keen on the survey) with those who skipped the open response questions (less keen on the survey).

As shown in Table 3, only number of adult cohabitants and age have significant differences, and in both cases, the effect size (η^2) is very small. This is consistent with minimal response bias.

Table 1 First principle components analysis*

Variable	Component			
	1	2	3	4
Δ P8	0.740			
Δ P2	0.715			
Δ P9	0.704	**0.304**		
Δ P6	0.699			
Δ P4	0.669			
Δ P3	0.645			
Δ P5	0.64			
Δ P1	0.563			
Δ P7	**0.356**			
Δ WP1		0.838		
Δ WP2		0.791		
Δ WP3		0.782		
Δ WP5		0.734		
Δ WP4		0.727		
Erg6			0.802	
Erg5			0.748	
Erg2			0.666	
Erg3			0.645	
Erg1	**0.306**		0.640	
Erg4			0.628	
DP3				0.688
DP1				0.661
DP5				0.568
DP2				0.565
DP4				0.493

*Rotation converged in 5 iterations. Coefficients < 0.3 suppressed

5.3 Demographics

Respondents were disproportionately male (81% vs. 18% female and 1% non-binary) and overwhelmingly employed full-time (94%) with a median age of 30–34. Participants were generally well-educated (Fig. 2). Most respondents (53%) live with one other adult, while 18% live with no other adults and the rest live with two or more people. 27% live with one or more children under 12. 8% indicate that they may have a disability that affects their work. Mean work experience is 9.3 years ($\sigma = 7.3$). Mean experience working from home is 1.3 years ($\sigma = 2.5$); however, 58% of respondents had no experience working from home before COVID-19.

Participants hail from 53 countries (Table 4) and organizations ranging from 0–9 employees to more than 100,000 (Fig. 3). Many participants listed multiple roles but 80% included software developer or equivalent among them, while the rest were other kinds of software professionals (e.g. project manager, quality assurance analyst).

Table 2 Second principle components analysis

Variable	Component			
	1	2	3	4
Δ P2	0.721			
Δ P8	0.718			
Δ P6	0.703			
Δ P4	0.679			
Δ P3	0.651			
Δ P5	0.649			
Δ P1	0.566			
Δ WB1		0.845		
Δ WB2		0.797		
Δ WB3		0.790		
Δ WB5		0.740		
Δ WB4		0.732		
Erg6			0.803	
Erg5			0.745	
Erg2			0.669	
Erg1			0.646	
Erg3			0.644	
Erg4			0.629	
DP3				0.685
DP1				0.666
DP2				0.570
DP5				0.565
DP4				0.490

Notes: Rotation convErged in 5 iterations; correlations < 0.3 suppressed

Table 3 Analysis of response bias (one-way ANOVA)

Variable	F	Sig.	η^2
age	4.250	0.039	0.002
disability	0.117	0.733	0.000
education	0.153	0.696	0.000
adultCohabitants	19.037	0.000	0.009
childCohabitants	0.358	0.550	0.000
experience	3.381	0.066	0.002
remoteExperience	0.013	0.910	0.000
organizationSize	0.330	0.566	0.000
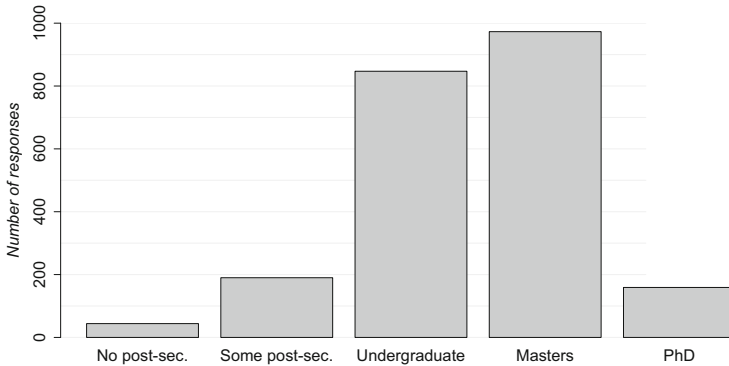
Fig. 2 Participants' education levels

Seven participants ($< 1\%$) tested positive for COVID-19 and six more ($< 1\%$) live with someone with COVID-19; 4% of respondents indicated that a close friend or family member had tested positive, and 13% were currently or recently quarantined.

5.4 Change in Wellbeing and Productivity

Hypothesis H1: Developers will have lower wellbeing while working from home due to COVID-19. Participants responded to the WHO-5 wellbeing scale twice—once referring to the 28-day period before switching to work from home and once referring to the period while working from home. We estimate wellbeing before and after by summing each set of five items, and then compare the resulting distributions (see Fig. 4). Since both scales deviate significantly from a normal distribution (Shapiro-Wilk test; $p < 0.001$; see Fig. 4), we compare the distributions using the two-sided paired Wilcoxon signed rank test. To estimate effect size, we use Cliff's delta (with 95% confidence level).

Hypothesis H1 is supported. (Wilcoxon signed rank test $V = 645610$; $p < 0.001$; $\delta = 0.12 \pm 0.03$).

Hypothesis H2: Developers will have lower perceived productivity while working from home due to COVID-19. Like the wellbeing scale, participants answered the HPQ

Table 4 Respondents' countries of residence

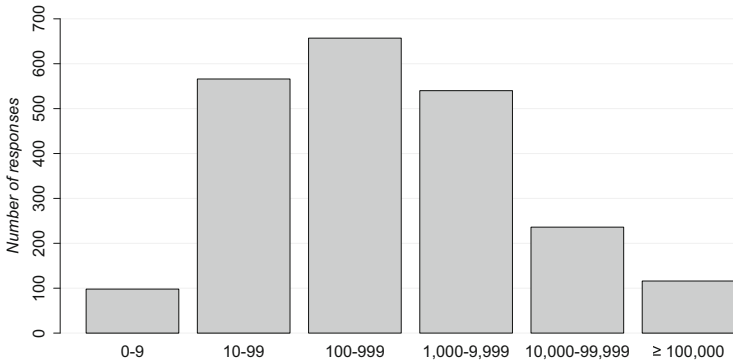| Country	n	%	Country	n	%
Germany	505	22.7%	Japan	53	2.4%
Russia	366	16.4%	Spain	52	2.3%
Brazil	272	12.2%	Iran	40	1.8%
Italy	173	7.8%	Austria	29	1.3%
United States	99	4.4%	Canada	27	1.2%
South Korea	81	3.6%	Switzerland	20	0.9%
Belgium	77	3.5%	United Kingdom	20	0.9%
China	76	3.4%	n/a	20	0.9%
Turkey	66	3.0%	Other	194	8.7%
India	55	2.5%			

Fig. 3 Organization sizes

productivity scale twice. Again, we estimate productivity before and after by summing each set of items (after correcting reversed items and omitting items 7 and 9; see Section 5.2). Again, the distributions are not normal (Shapiro-Wilk test; $p < 0.001$; see Fig. 4), so we use the Wilcoxon signed rank test and Cliff's delta.

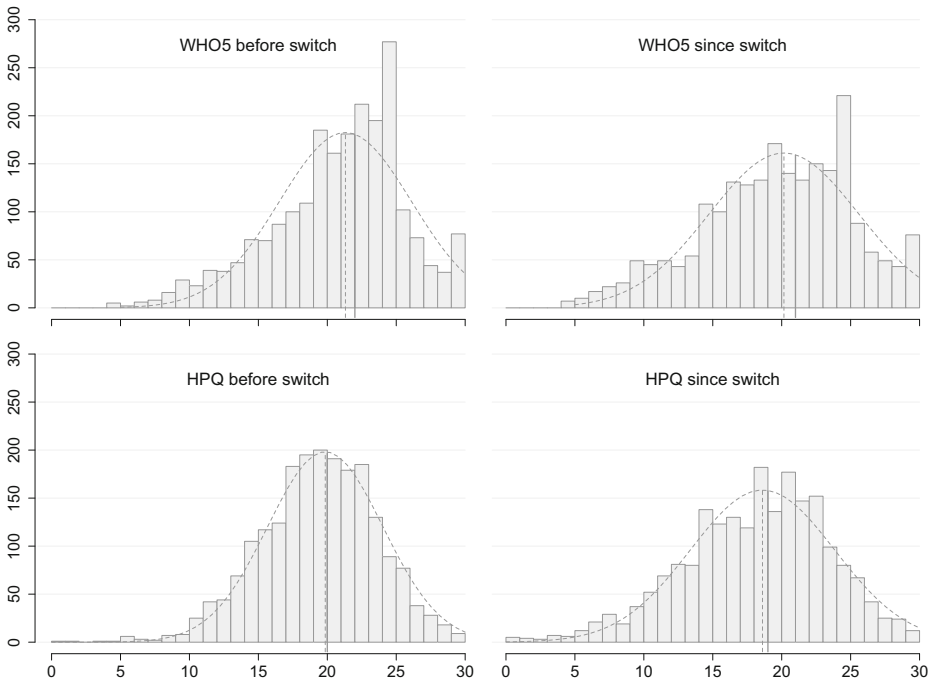**Hypothesis H2 is supported**. (Wilcoxon signed rank test $V = 566520$; $p < 0.001$; $\delta = 0.13 \pm 0.03$).

Fig. 4 Distribution of ratings on the WHO-5 and HPQ scales before and since switching to working form home with mean (dashed line) and median (solid line) values (2,194 complete cases for WHO-5 and 2,078 for HPQ)

5.5 Structural Equation Model

To test our remaining hypotheses, we use structural equation modeling (SEM). Briefly, SEM is used to test theories involving constructs (also called latent variables). A *construct* is a quantity that cannot be measured directly (Ralph and Tempero 2018). Fear, disaster preparedness, home office ergonomics, wellbeing and productivity are all constructs. In contrast, age, country, and number of children are all directly measurable.

To design a structural equation model, we first define a *measurement model*, which maps each *reflective indicator* into its corresponding construct. For example, each of the five items comprising the WHO-5 wellbeing scale is modeled as a reflective indicator of wellbeing. SEM uses confirmatory factor analysis to estimate each construct as the shared variance of its respective indicators.

Next, we define the *structural model*, which defines the expected relationships among the constructs. The constructs we are attempting to predict are referred to as *endogenous*, while the predictors are *exogenous*.

SEM uses a path modeling technique (e.g. regression) to build a model that predicts the endogenous (latent) variables based on the exogenous variables, and to estimate both the strength of each relationship and the overall accuracy of the model.[4]

As mentioned, the first step in a SEM analysis is to conduct a confirmatory factor analysis to verify that the measurement model is consistent (Table 5). Here, the latent concepts Ergonomics and DisasterPreparedness are captured by their respective exogenous variables. Fear is not included because it is computed manually (see Section 4.2). ΔWellbeing is the difference in a participant's emotional wellbeing before and after switching to working from home. This latent concept is captured by five exogenous variables, $\Delta WB1, \ldots, \Delta WB5$. Similarly, ΔProductivity represents the difference in perceived productivity, before and after switching to working from home.

The confirmatory factor analysis converged (not converging would suggest a problem with the measurement model) and all of the indicators load well on their constructs. The lowest estimate, 0.716 for DP2, is still quite good. The estimates for ΔP2 through ΔP6 are negative because these items were reversed (i.e. higher score = worse productivity). Note that factor loadings greater than one do not indicate a problem because they are regression coefficients, not correlations (Jöreskog 1999).

Having reached confidence in our measurement model, we construct our structural model by representing all of the hypotheses stated in Section 3 as regressions (e.g. ΔWellbeing $\sim$ DisasterPreparedness + Fear + Ergonomics).

In principle, we use all control variables as predictors for all latent variables. In practice, however, this leads to too many relationships and prevents the model from converging. Therefore, we evaluate the predictive power of each control variable, one at a time, and include it in a regression only where it makes at least a marginally significant ($p < 0.1$) difference. Here, using a higher than normal p-value is more conservative because we are dropping predictors rather than testing hypotheses. Country of residence and language of questionnaire are not included because SEM does not respond well to nominal categorical variables (see Section 5.6).

---

[4]Data was analyzed using the R package lavaan 0.6-5 available at http://lavaan.ugent.be/.

Table 5 Confirmatory factor analysis

Construct	Indicator	Estimate	Std.Err	z-value	$P(>|z|)$
ΔWellbeing =~	ΔWB1	1.000			
	ΔWB2	0.896	0.016	54.518	0
	ΔWB3	0.955	0.016	58.917	0
	ΔWB4	0.804	0.018	44.686	0
	ΔWB5	0.848	0.017	51.041	0
ΔProductivity =~	ΔP1	1.000			
	ΔP2	−1.268	0.053	−24.084	0
	ΔP3	−1.120	0.053	−20.979	0
	ΔP4	−1.239	0.053	−23.263	0
	ΔP5	−1.229	0.055	−22.266	0
	ΔP6	−1.306	0.058	−22.677	0
	ΔP8	1.460	0.057	25.512	0
Ergonomics =~	Erg1	1.000			
	Erg2	0.964	0.035	27.395	0
	Erg3	0.820	0.037	22.128	0
	Erg4	0.937	0.035	26.663	0
	Erg5	1.064	0.034	31.535	0
	Erg6	1.258	0.035	35.821	0
Disaster	DP1	1.000			
Preparedness =~	DP2	0.716	0.089	8.079	0
	DP3	1.181	0.112	10.521	0
	DP4	0.923	0.105	8.805	0
	DP5	1.186	0.120	9.888	0

Notes: converged after 50 iterations with 185 free parameters ($n = 1377$); estimates may exceed 1.0 because they are regression coefficients, not correlations as in Principal Component Analysis; negative estimates indicate reversed items

Since the exogenous variables are ordinal, the weighted least square mean variance (WLSMV) estimator was used. WLSMV uses diagonally weighted least squares to estimate the model parameters, but it will use the full weight matrix to compute robust standard errors, and a mean- and variance-adjusted test statistic. In short, the WLSMV is a robust estimator which does not assume a normal distribution, and provides the best option for modelling ordinal data in SEM (Brown 2006). We use the default Nonlinear Minimization subject to Box Constraints (NLMinB) optimizer.

For missing data, we use pairwise deletion: we only keep those observations for which both values are observed (this may change from pair to pair). By default, since we are also dealing with categorical exogenous variables, the model is set to be conditional on the exogenous variables.

The model was executed and all diagnostics passed, that is, lavaan ended normally after 97 iterations with 212 free parameters and $n = 1377$. We evaluate model fit by inspecting several indicators (cf. Hu and Bentler 1999, for cut-offs):

– The Comparative Fit Index (CFI = 0.961) and Tucker-Lewis Index (TLI = 0.979), which compare the model's fit against the null model, should be at least 0.95.
– The Root Mean Square Error of Approximation (RMSEA = 0.051, 90% CI [0.048; 0.053]) should be less than 0.06.
– The Standardized Root Mean Square Residual (SRMR = 0.067) should be less than 0.08 (for large sample sizes).

In summary, all diagnostics indicate the model is safe to interpret (i.e. CFI = 0.961, RMSEA = 0.051, SRMR = 0.067).

Figure 5 illustrates the supported structural equation model. The arrows between the constructs show the supported causal relationships. The path coefficients (the numbers on the arrows) indicate the relative strength and direction of the relationships. For example, the arrow from Disaster Preparedness to Fear indicates that the hypothesis that Disaster Preparedness affects Fear was supported. The label (-0.336) indicates an inverse relationship (more Disaster Preparedness leads to less Fear) and 0.336 indicates the strength of the relationship.

> Based on this model, **Hypotheses H1–H3, H5, H6, and H8–H10 are supported; Hypotheses H4 and H7 are not supported**. That is, change in wellbeing and change in perceived productivity are directly related; change in perceived productivity depends on home office ergonomics and disaster preparedness; change in wellbeing depends on ergonomics and fear; and disaster preparedness is inversely related to fear.

5.6 Exploratory Findings

Inspecting the detailed SEM results (Table 6) reveals numerous interesting patterns. Direct effects include:

– People who live with small children have significantly less ergonomic home offices. This is not surprising because the ergonomics scale included items related to noise and distractions.
– Women tend to be more fearful. This is consistent with studies on the SARS epidemic, which found that women tended to perceive the risk as higher (Brug et al. 2004).
– People with disabilities are less prepared for disasters, have less ergonomic offices and are more afraid.
– People who live with other adults are more prepared for disasters.
– People who live alone have more ergonomic home offices.
– People who have COVID-19 or have family members, housemates or close friends with COVID-19 tend to be more afraid, more prepared, and have worse wellbeing since working from home.
– People who are more isolated (i.e. not leaving home at all, or only for necessities) tend to be more afraid.
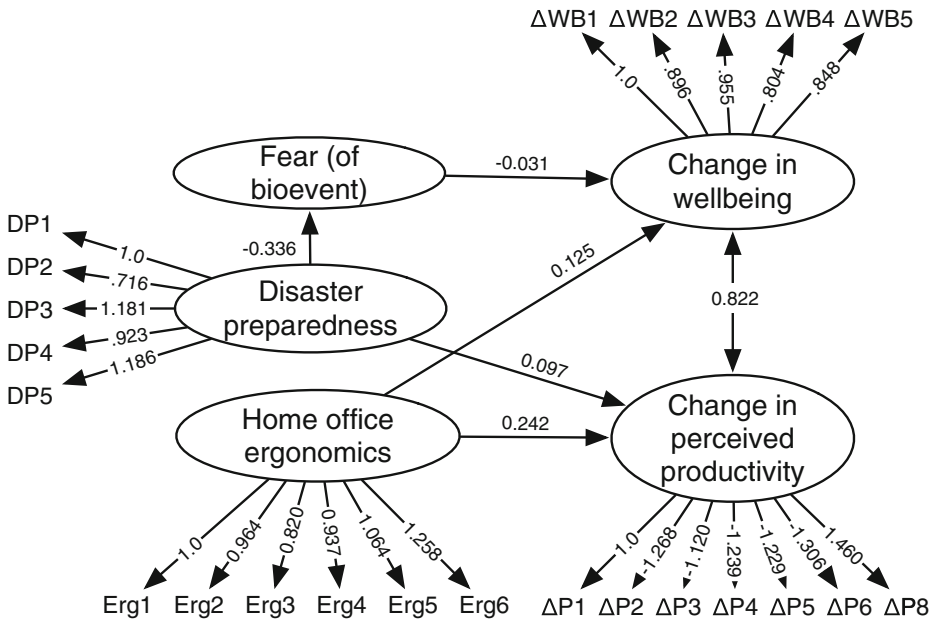
Fig. 5 Supported model of developer wellbeing and productivity. Note: error terms, unsupported hypotheses and control variables are omitted for clarity

Some indirect effects are also apparent, but are more difficult to interpret. For example, changes in productivity and wellbeing are closely related. Hypothesis H4 may be unsupported because change in productivity is mediating the effect of disaster preparedness on change in wellbeing. Similarly, Hypothesis H7 may not be unsupported because change in wellbeing is mediating the relationship between fear and change in productivity. Furthermore, control variables including gender, children and disability may have significant effects on wellbeing or productivity that are not obvious because they are mediated by another construct. Some variables have conflicting effects. For example, disability has not only a direct positive effect on productivity but also an indirect negative effect (through fear). So, is the pandemic harder on people with disabilities? More research is needed to explore these relationships.

Above, we mentioned omitting language and country because SEM does not respond well to nominal categorical variables. We tried anyway, and both language and country were significant predictors for all latent variables, but, as expected, including so many binary dummy variables makes the model impossible to interpret. While our analysis suggests that country, language (and probably culture) have significant effects on disaster preparedness, ergonomics, fear, wellbeing and productivity, more research is need to understand the nature of these effects (see Section 6.3).

5.7 Organizational Support

Table 7 shows participants' opinions of the helpfulness of numerous ways their organizations could support them. Several interesting patterns stand out from this data:

Table 6 Structural equation model regressions

Construct	Predictor	Estimate	Std.Err	z-value	$P(> |z|)$
Disaster	adultCohabitants	0.080	0.019	4.234	0.000
Preparedness ∼	disability	−0.179	0.059	−3.035	0.002
	covidStatus	0.073	0.032	2.260	0.024
	education	−0.050	0.026	−1.882	0.060
Ergonomics ∼	children	−0.163	0.031	−5.184	0.000
	adultCohabitants	−0.047	0.019	−2.457	0.014
	disability	−0.110	0.057	−1.932	0.053
	remoteExperience	0.044	0.026	1.709	0.087
Fear ∼	isolation	0.502	0.105	4.764	0.000
	DisasterPreparedness	−0.336	0.106	−3.161	0.002
	role	−0.356	0.116	−3.056	0.002
	covidStatus	0.196	0.075	2.607	0.009
	gender	0.273	0.122	2.241	0.025
	disability	0.265	0.119	2.227	0.026
	education	−0.122	0.060	−2.047	0.041
	children	0.116	0.063	1.831	0.067
ΔWellbeing ∼	Ergonomics	0.125	0.033	3.813	0.000
	covidStatus	−0.121	0.040	−3.041	0.002
	Fear	−0.031	0.012	−2.542	0.011
	age	0.097	0.044	2.204	0.028
	DisasterPreparedness	−0.020	0.049	−0.416	0.678
ΔProductivity ∼	Ergonomics	0.242	0.024	10.233	0.000
	DisasterPreparedness	0.097	0.035	2.788	0.005
	adultCohabitants	0.041	0.015	2.752	0.006
	disability	0.124	0.049	2.513	0.012
	age	0.070	0.032	2.220	0.026
	Fear	−0.002	0.009	−0.204	0.838
ΔWellbeing ∼	ΔPerformance	0.822	0.045	18.361	0.000

Notes: converged after 97 iterations; Latent variables capitalized (e.g. Fear); direct measurements in camelCase (e.g. age, adultCohabitants)

- Only action #1—paying developer's home internet charges—is perceived as helpful by more than half of participants and less than 10% of companies appear to be doing that.
- The action most companies are taking (#20, having regular meetings) is not perceived as helpful by most participants.
- There appears to be no correlation between actions developers believe would help and actions employers are actually taking.
- There is little consensus among developers about what their organizations should do to help them.

Table 7 Organizational support actions in order of perceived helpfulness*

#	Action	Helpful	Following
1	My organization will pay for some or all of my internet charges	51.9%	9.8%
2	My organization will buy new equipment we need to work from home	49.2%	30.9%
3	My organization is encouraging staff to use this time for professional training	47.7%	24.3%
4	My organization has reassured me that they understand if my work performance suffers	47.4%	40.5%
5	My organization is providing activities to occupy staff member's children	46.4%	7.2%
6	My organization is sending food to staff working from home	44.5%	4.0%
7	My organization is providing at-home exercise programs	41.4%	15.8%
8	My organization has reassured me that I will keep my job	40.2%	62.4%
9	My organization has reassured me that I can take time off if I'm sick or need to care for dependents	40.1%	65.5%
10	My organization is improving documentation of its processes (e.g. how code changes are approved)	37.4%	34.7%
11	My organization will pay for software we need to work from home	36.8%	54.7%
12	My team is peer reviewing commits, change requests or pull requests (peer code review)	36.5%	63.1%
13	I can (or could) take equipment (e.g. monitors) home from my workplace	36.0%	73.9%
14	My organization has reassured me that I will continue to be paid	34.7%	75.2%
15	My team uses a build system to automate compilation and testing	34.3%	62.9%
16	Someone is keeping high priority work ready and our backlog organized	33.1%	60.0%
17	My team has good work-from-home infrastructure (e.g. source control, VPN, remote desktop, file sharing)	32.6%	86.4%
18	My team is having virtual social events (e.g. via video chat)	32.1%	56.1%
19	My organization is encouraging staff to touch base regularly with each other	30.8%	62.4%
20	My team is continuing to have regular meetings (e.g. via video chat)	28.5%	88.9%
21	My team is avoiding synchronous communication (e.g. video chat)	25.5%	14.3%
22	For most of the day, I work with an open video or audio call to some or all of my team	23.3%	26.7%

*number of respondents who indicated that this practice is or would be *helpful* and number of respondents who indicated that their organizations are *following* this recommendation (n=2225)

In hindsight, the structure of this question may have undermined discrimination between items. Future work could investigate a better selection of actions, and possibly ask participants for their "top N" items to improve reliability. Moreover, the helpfulness of actions may depend on where the participant lives; for example, in countries with a weaker social safety net, reassuring employees that they will keep their jobs, pay and benefits may be more important.

5.8 Summary Interpretation

This study shows that software professionals who are working from home during the pandemic are experiencing diminished emotional wellbeing and productivity, which are closely related. Furthermore, poor disaster preparedness, fear related to the pandemic, and poor home office ergonomics are exacerbating this reduction in wellbeing and productivity. Moreover, women, parents and people with disabilities may be disproportionately affected. In addition, dissensus regarding what organizations can do to help suggests that no single action is universally helpful; rather, different people need different kinds of support.

6 Discussion

6.1 Recommendations

Organizations need to accept that expecting normal productivity under these circumstances is unrealistic. Pressuring employees to maintain normal productivity will likely make matters worse. Furthermore, companies should avoid making any decisions (e.g. layoffs, promotions, bonuses) based on productivity during the pandemic because any such decision may be prejudiced against protected groups. If a member of a protected group feels discriminated against due to low productivity at this time, we recommend contacting your local human rights commission or equivalent organization.

Because productivity and wellbeing are so closely related, the best way to improve productivity is to help employees maintain their emotional wellbeing. However, no single action appears beneficial to everyone, so organizations should talk to their employees to determine what they need.

Helping employees improve the ergonomics of their work spaces, in particular, should help. However, micromanaging foot positions, armrests and hip angles is not what we mean by ergonomics. Rather, companies should ask broad questions such as "what do you need to limit distractions and be more comfortable?" Shipping an employee a new office chair or noise cancelling headphones could help significantly.

Meanwhile, professionals should try to accept that their productivity may be lower and stop stressing about it. Similarly, professionals should try to remember that different people are experiencing the pandemic in very different ways—some people may be more productive than normal while others struggle to complete any work through no fault of their own. It is crucial to support each other and avoid inciting conflict over who is working harder.

6.2 Limitations and Threats to Validity

The above recommendations should be considered in the context of the study's limitations.

Sampling bias Random sampling of software developers is rare (Amir and Ralph 2018) because there are no lists of all the software developers, projects, teams or organizations in

the world or particular jurisdictions (Baltes and Ralph 2020). We therefore combined convenience and snowball sampling with a strategy of finding a co-author with local knowledge to translate, localize and advertise the questionnaire in a locally effective way. On one hand, the convenience/snowball strategy may bias the sample in unknown ways. On the other hand, our translation and localization strategy demonstrably increased sample diversity, leading to one of the largest and broadest samples of developers ever studied, possible due to a large, international and diverse research team. Any random sample of English-speaking developers is comparatively ethnocentric. The sample is not balanced; for instance, many more respondents live in Germany than all of southeast Asia, but we attempt to correct for that (see *Internal Validity*, below).

Response Bias Meanwhile, we found minimal evidence of response bias (in Section 5.2). However, because the questionnaire is anonymous and Google Forms does not record incomplete responses, response bias can only be estimated in a limited way. Someone could have taken the survey more than once or entered fake data. Additionally, large responses from within a single country could skew the data but we correct for company size, language and numerous demographic variables to mitigate this.

Construct validity To enhance construct validity, we used validated scales for wellbeing, productivity, disaster preparedness and fear/resilience. Post-hoc construct validity analysis suggests that all four scales, as well as the ergonomics scale we created, are sound (Section 5.2). However, perceived productivity is not the same as actual productivity. Although the HPQ scale correlates well with objective performance data in other fields (Kessler et al. 2003), it may not in software development or during pandemics. Similarly, we asked respondents their opinion of numerous potential mechanisms for organizational support. Just because companies are taking some action (e.g. having regular meetings) or respondents believe in the helpfulness of some action (e.g. paying their internet bills), does not mean that those actions will actually improve productivity or wellbeing.

Measurement Validity. There is much debate about whether 5- and 6-point responses should be treated as ordinal or interval. CFA and SEM are often used with these kinds of variables in social sciences despite assuming at least interval data. Some evidence suggests that CFA is robust against moderate deviations from normality, including arguably-ordinal questionnaire items (cf. Flora and Curran 2004). We tend not to worry about treating data as interval as long as, in principle, the data is drawn from a continuous distribution. Additionally, due to a manual error, the Italian version was missing organizational support item 11: "My team uses a build system to automate compilation and testing." The survey may therefore under-count the frequency and importance of this item by up to 10%.

Conclusion validity. We use structural equation modeling to fit a theoretical model to the data. Indicators of model fit suggest that the model is sound. Moreover, SEM enhances conclusion validity by correcting for multiple comparisons, measurement error (by inferring latent variables based on observable variables), testing the entire model as a whole (instead of one hypothesis at a time) and controlling for extraneous variables (e.g. age, organization size). SEM is considered superior to alternative path modeling techniques such as partial least squares path modeling (Rönkkö and Evermann 2013). While a Bayesian approach might have higher conclusion validity (Furia et al. 2019), none of the Bayesian SEM tools (e.g. Blaavan) we are aware of support ordered categorical variables. The main remaining threat to conclusion validity is overfitting the structural model. More research is needed to determine whether the model overstates any of the supported effects.

Internal validity. To infer causality, we must demonstrate correlation, precedence and the absence of third variable explanations. SEM demonstrates correlation. SEM does not demonstrate precedence; however, we can be more confident in causality where precedence only makes sense in one direction. For example, having COVID-19 may reduce one's productivity, but feeling unproductive cannot give someone a specific virus. Similarly, it seems more likely that a more ergonomic office might make you more productive than that being more productive leads to a more ergnomic office. Meanwhile, we statistically controlled for numerous extraneous variables (e.g. age, gender, education level, organization size). However, other third-variable explanations cannot be discounted. Developers who work more overtime, for example, might have lower wellbeing, worse home office ergonomics, and reduced disaster preparedness. Other confounding variables might include individual differences (e.g. personality), team dynamics, organizational culture, family conflict, past medical history and wealth.

6.3 Implications for Researchers and Future Work

For researchers, this paper opens a new research area intersecting software engineering and crisis, disaster and emergency management. Although many studies explore remote work and distributed teams, we still need a better understanding of how stress, distraction and family commitments affect developers working from home *during crises, bioevents and disasters*. More research is needed on how these events affect team dynamics, cohesion, performance, as well as software development processes and practices.

More specifically, the dataset we publish alongside this paper can be significantly extended. Abundant quantitative data is available regarding different countries, and how those countries reacted to the COVID-19 pandemic. Country data could be merged with our dataset to investigate how different contexts, cultures and political actions affect developers. For example, the quality of a country's social safety net may affect fear.

Furthermore, the crisis continues. More longitudinal research is needed to understand its long-term effects on software professionals (e.g. burnout), projects (e.g. decreased velocity) and communities (e.g. trust issues). Research is also needed to understand how the crisis affects different kinds of professions. We focus on software developers because that is who software engineering research is responsible for, in the same way nursing researchers will study nurses and management researchers will study managers. Only comparing studies of different groups will reveal the extent to which our findings are specific to software professionals or generalizable to other knowledge workers.

This study does not investigate typical software engineering practices (e.g. pair programming, mutation testing) or debates (e.g. agile methods vs. model-driven engineering) because we do not believe that a team's software development methodology is a key antecedent of pandemic-induced changes to productivity and wellbeing. Further research is needed to confirm or refute our intuition.

6.4 Lessons Learned

This study taught us two valuable lessons about research methodology. First, collaborating with a large, diverse, international research team and releasing a questionnaire in multiple languages *with location-specific advertising* can generate a large, diverse, international sample of participants.

Second, Google Forms should not be used to conduct scientific questionnaire surveys. It is blocked in some countries. It does not record partial responses or bounce rates, hindering

analysis of response bias. URL parameter passing, which is typically used to determine how the respondent found out about the survey, is difficult. Exporting the data in different ways gives different variable orders, encouraging mistakes. Responses are recorded as (sometimes long) strings instead of numbers, overcomplicating data analysis. We should have used a research focused survey tool such as LimeSurvey(.org) or Qualtrics(.com).

7 Conclusion

The COVID-19 pandemic has created unique conditions for many software developers. Stress, isolation, travel restrictions, business closures and the absence of educational, child care and fitness facilities are all taking a toll. Working from home under these conditions is fundamentally different from normal working from home. This paper reports the first large-scale study of how working from home during a pandemic affects software developers. It makes several key contributions:

– evidence that productivity and wellbeing have declined;
– evidence that productivity and wellbeing are closely related;
– a model that explains and predicts the effects of the pandemic on productivity and wellbeing;
– some indication that different people need different kinds of support from their organizations (there is no silver bullet here);
– some indication that the pandemic may disproportionately affect women, parents and people with disabilities.

Furthermore, this study is exceptional in several ways: (1) the questionnaire used previously validated scales, which we re-validated using both principal components analysis and confirmatory factor analysis; (2) the questionnaire attracted an unusually large sample of 2225 responses; (3) the questionnaire ran in 12 languages, mitigating cultural biases; (4) the data was analyzed using highly sophisticated methods (i.e. structural equation modelling), which rarely have been utilized in software engineering research; (5) the study investigates an emerging phenomenon, providing timely advice for organizations and professionals; (6) the study incorporates research on emergency and disaster management, which is rarely considered in software engineering studies.

We hope that this study inspires more research on how software development is affected by crises, pandemics, lockdowns and other adverse conditions. As the climate crisis unfolds, more research intersecting disaster management and software engineering will be needed.

8 Data Availability

A comprehensive replication package including our (anonymous) dataset, instruments and analysis scripts is stored in the Zonodo open data archive at https://zenodo.org/record/3783511.

References

Amir B, Ralph P (2018) There is no random sampling in software engineering research. In: Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings, pp 344–345. https://doi.org/10.1145/3183440.3195001

Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD (2020) How will country-based mitigation measures influence the course of the covid-19 epidemic? The Lancet 395(10228):931–934

ARI (2020) Worry, gratitude & boredom: As covid-19 affects mental, financial health, who fares better; who is worse? http://angusreid.org/covid19-mental-health/

Bailey DE, Kurland NB (2002) A review of telework research: findings, new directions, and lessons for the study of modern work. J Organ Behav Int Ind Occup Organ Psychol Behav 23(4):383–400

Baker E, Avery GC, Crawford J (2007) Satisfaction and perceived productivity when professionals work from home. Res Pract Hum Resour Manag 15(1):37–62

Baltes S, Diehl S (2018) Towards a theory of software development expertise. In: Leavens GT, Garcia A, Pasareanu CS (eds) Proceedings of the 2018 ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/SIGSOFT FSE 2018. ACM Lake Buena Vista, pp 187–200. https://doi.org/10.1145/3236024.3236061

Baltes S, Ralph P (2020) Sampling in software engineering research: A critical review and guidelines. arXiv:2002.07764

Bao L, Li T, Xia X, Zhu K, Li H, Yang X (2020) How does working from home affect developer productivity? a case study of baidu during covid-19 pandemic. arXiv:2005.13167

Baruch Y (1996) Self performance appraisal vs direct-manager appraisal: a case of congruence. J Manag Psychol 11(6):50–65. https://doi.org/10.1108/02683949610129758

Baruch Y (2000) Teleworking: benefits and pitfalls as perceived by professionals and managers. Technol Work Employment 15(1):34–49

Blake KD, Blendon RJ, Viswanath K (2010) Employment and compliance with pandemic influenza mitigation recommendations. Emerg Infect Diseases 16(2):212

Bracha HS, Burkle FM (2006) Utility of fear severity and individual resilience scoring as a surge capacity, triage management tool during large-scale, bio-event disasters. Prehospital Disast Med 21(5):290–296

Brown TA (2006) Confirmatory factor analysis for applied research. Methodology in the social sciences, Guilford Publications

Brug J, Aro AR, Oenema A, De Zwart O, Richardus JH, Bishop GD (2004) Sars risk perception, knowledge, precautions, and information sources, the netherlands. Emerg Infect Diseases 10(8):1486

Cascio WF (2000) Managing a virtual workplace. Acad Manag Perspect 14(3):81–90

Chrystal KA, Mizen PD (2003) Goodhart's law: Its origins, meaning and implications for monetary policy. Central Bank Monetary Theory Pract Essays Honour of Charles Goodhart 1:221–243

Dall'Ora C, Ball J, Recio-Saucedo A, Griffiths P (2016) Characteristics of shift work and their impact on employee performance and wellbeing: a literature review. Int J Nurs Stud 57:12–27

Davenport TH, Pearlson K (1998) Two cheers for the virtual office. MIT Sloan Manag Rev 39(4):51

DiGiovanni C, Conley J, Chiu D, Zaborski J (2004) Factors influencing compliance with quarantine in toronto during the 2003 sars outbreak. Biosecur Bioterrorism Biodefense Strat Pract Sci 2(4):265–272

Donnelly N, Proctor-Thomson SB (2015) Disrupted work: home-based teleworking (hbtw) in the aftermath of a natural disaster. Technol Work Employment 30(1):47–61

Duffy C (2020) Big tech firms ramp up remote working orders to prevent coronavirus spread. https://www.cnn.com/2020/03/10/tech/google-work-from-home-coronavirus/index.html

Duxbury L, Higgins C, Neufeld D (1998) Telework and the balance between work and family: Is telework part of the problem or part of the solution? In: Igbaria M, Tan M (eds) The virtual workplace. Hershey

Evers KE, Castle PH, Prochaska JO, Prochaska JM (2014) Examining relationships between multiple health risk behaviors, well-being, and productivity. Psychol Rep 114(3):843–853

Fisher JR, Tran TD, Hammarberg K, Sastry J, Nguyen H, Rowe H, Popplestone S, Stocker R, Stubber C, Kirkman M et al (2020) Mental health of people in Australia in the first month of COVID-19 restrictions: a national survey. The Medical Journal of Australia. https://www.mja.com.au/journal/2020/mental-health-people-australia-first-month-covid-19-restrictions-restrictionsnational-survey

Flora DB, Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 9(4):466

Forsgren N (2020) Octoverse spotlight: An analysis of developer productivity, work cadence, and collaboration in the early days of covid-19. https://github.blog/2020-05-06-octoverse-spotlight-an-analysis-of-developer-productivity-work-cadence-and-collaboration-in-the-early-days-of-covid-19/

Furia CA, Feldt R, Torkar R (2019) Bayesian data analysis in empirical software engineering research. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2019.2935974

Goodhart CAE (1984) Monetary Theory and Practice: The UK Experience. Macmillan Press

Graziotin D, Wang X, Abrahamsson P (2015) How do you feel, developer? an explanatory theory of the impact of affects on programming performance. PeerJ Comput Sci 1:e18

Graziotin D, Fagerholm F, Wang X, Abrahamsson P (2017) Consequences of unhappiness while developing software. In: 2017 IEEE/ACM 2Nd international workshop on emotion awareness in software engineering (SEmotion). IEEE, pp 42–47

Graziotin D, Fagerholm F (2019) Happiness and the productivity of software engineers. In: Sadowski C, Zimmermann T (eds) Apress open / Springer, pp 109–124. https://doi.org/10.1007/978-1-4842-4221-6_10

Hanselman S (2020) Tweet. https://twitter.com/shanselman/status/1252040170783641600

Harper C, Satchell L, Fido D, Latzman R (2020) Functional fear predicts public health compliance in the COVID-19 pandemic. International Journal of Mental Health and Addiction, pp 1–21. https://doi.org/10.1007/s11469-020-00281-5

Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 6(1):1–55. https://doi.org/10.1080/10705519909540118

Hyman J, Baldry C, Scholarios D, Bunzel D (2003) Work–life imbalance in call centres and software development. Br J Ind Relat 41(2):215–239

Inalhan G, Ng CF (2010) Teleworker's home office: an extension of corporate office? Facilities 28(3/4):137–155

Jaspan C, Sadowski C (2019) No single metric captures productivity. In: Sadowski C, Zimmermann T (eds) Rethinking Productivity in Software Engineering. Apress open / Springer, pp 13–20. https://doi.org/10.1007/978-1-4842-4221-6_2

Jöreskog KG (1999) How large can a standardized coefficient be. http://www.statmodel.com/download/Joreskog.pdf

Kaplan J, Frias L, McFall-Johnsen M (2020) A third of the global population is on coronavirus lockdown — here's our constantly updated list of countries and restrictions. https://www.businessinsider.com/countries-on-lockdown-coronavirus-italy-2020-3?op=1

Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, Pronk N, Simon G, Stang P, Ustun TB et al (2003) The world health organization health and work performance questionnaire (HPQ). J Occup Environ Med 45(2):156–174

Kim Y, Zhong W, Jehn M, Walsh L (2015) Public risk perceptions and preventive behaviors during the 2009 h1n1 influenza pandemic. Disaster Med Publ Health Prepared 9(2):145–154

Ko AJ (2019) Why we should not measure productivity. In: Sadowski C, Zimmermann T (eds) Rethinking Productivity in Software Engineering. Apress open / Springer, pp 21–26. https://doi.org/10.1007/978-1-4842-4221-6_3

Lavelle J (2020) Gartner CFO survey reveals 74% intend to shift some employees to remote work permanently. https://www.gartner.com/en/newsroom/press-releases/2020-04-03-gartner-cfo-surey-reveals-74-percent-of-organizations-to-shift-some-employees-to-remote-work-permanently2

Loose VW, Vargas VN, Warren DE, Starks SJ, Brown TJ, Smith BJ (2010) Economic and policy implications of pandemic influenza. Report SAND2010-1910 Sandia National Laboratories. Albuquerque, New Mexico March

Madhav N, Oppenheim B, Gallivan M, Mulembakani P, Rubin E, Wolfe N (2017) Pandemics: risks, impacts, and mitigation. In: Jamison, D. et al. (eds.) Disease Control Priorities, ch. 17. https://doi.org/10.1596/978-1-4648-0527-1_ch17

McInerney CR (1999) Working in the virtual office: Providing information and knowledge to remote workers. Library Inf Sci Res 21(1):69–89

McLeod L, MacDonell SG (2011) Factors that affect software systems development project outcomes: a survey of research. ACM Comput Surv (CSUR) 43(4):1–56

Meyer AN, Murphy GC, Zimmermann T, Fritz T (2017) Retrospecting on work and productivity: A study on self-monitoring software developers' work. PACMHCI 1(CSCW):79:1–79:24. https://doi.org/10.1145/3134714

North CS, Pfefferbaum B, Hong BA, Gordon MR, Kim YS, Lind L, Pollio DE (2010) The business of healing: Focus group discussions of readjustment to the post-9/11 work environment among employees of affected agencies. J Occup Environ Med 52(7):713–718

Paton D (2008) Community resilience: Integrating individual, community and societal perspectives. In: Gow K, Paton D (eds) The phoenix of natural disasters, Community resilience. Nova Science Publishers, New York, pp 13–31

Pérez MP, Sánchez AM, de Luis Carnicer P, Jiménez MJV (2004) A technology acceptance model of innovation adoption: the case of teleworking. Eur J Innov Manag 7(4):280–291. https://doi.org/10.1108/14601060410565038

Perry SJ, Rubino C, Hunter EM (2018) Stress in remote work: two studies testing the demand-control-person model. Eur J Work Organ Psychol 27(5):577–593

Prati G, Pietrantoni L, Zani B (2011) A social-cognitive model of pandemic influenza h1n1 risk perception and recommended behaviors in italy. Risk Anal Int J 31(4):645–656

Ralph P, Tempero E (2018) Construct validity in software engineering research and software metrics. In: Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018, pp 13–23. https://doi.org/10.1145/3210459.3210461

Rebmann T (2009) Infectious disease disasters: Bioterrorism, emerging infections, and pandemics. In: APIC text of infection control and epidemiology, 3rd edn. Association for Professionals in Infection Control and Epidemiology, Inc., pp 120.1–120.22

Roberto AJ, Goodall CE, Witte K (2009) Raising the alarm and calming fears: Perceived threat and efficacy during risk and crisis. In: Handbook of risk and crisis communication. Routledge, pp 287–303

Ronan KR, Alisic E, Towers B, Johnson VA, Johnston DM (2015) Disaster preparedness for children and families: a critical review. Curr Psych Rep 17(7):58

Rönkkö M, Evermann J (2013) A critical examination of common beliefs about partial least squares path modeling. Organ Res Methods 16(3):425–448

Russell H, O'Connell PJ, McGinnity F (2009) The impact of flexible working arrangements on work–life conflict and work pressure in Ireland. Gender Work Organ 16(1):73–97

Saldaña J (2015) The coding manual for qualitative researchers. Sage

Shultz JM, Cooper JL, Baingana F, Oquendo MA, Espinel Z, Althouse BM, Marcelin LH, Towers S, Espinola M, McCoy CB et al (2016) The role of fear-related behaviors in the 2013–2016 west Africa ebola virus disease outbreak. Curr Psych Rep 18(11):104

Spataro J (2020) Helping our developers stay productive while working remotely. https://www.microsoft.com/en-us/microsoft-365/blog/2020/03/20/helping-developers-stay-productive-working-remotely/

Storey MA, Zimmermann T, Bird C, Czerwonka J, Murphy B, Kalliamvakou E (2019) Towards a theory of software developer job satisfaction and perceived productivity. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2019.2944354

Taha S, Matheson K, Cronin T, Anisman H (2014) Intolerance of uncertainty, appraisals, coping, and anxiety: The case of the 2009 h1n1 pandemic. Brit J Health Psychol 19(3):592–605

Teasdale E, Yardley L, Schlotz W, Michie S (2012) The importance of coping appraisal in behavioural responses to pandemic flu. Brit J Health Psychol 17(1):44–59

Thommes E, Cojocaru M, Athar S (2016) Absenteeism impact on local economy during a pandemic via hybrid sir dynamics. In: International Conference on Dynamics of Disasters, Springer, pp 309–328

Topp CW, Østergaard SD, Søndergaard S, Bech P (2015) The WHO-5 well-being index: a systematic review of the literature. Psych Psychosomat 84(3):167–176

WHO (2020a) Coronavirus disease 2019 (COVID-19): situation report 70. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200430-sitrep-101-covid-19.pdf?sfvrsn=2ba4e093_2

WHO (2020b) Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov). https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)

Wrobel MR (2013) Emotions in the software development process. In: 6th international conference on human system interactions (HSI). IEEE, pp 518–523

Xiang YT, Zhao YJ, Liu ZH, Li XH, Zhao N, Cheung T, Ng CH (2020) The COVID-19 outbreak and psychiatric hospitals in china: managing challenges through mental health service reform. Int J Biol Sci 16(10):1741

Yong AG, Lemyre L, Pinsent C, Krewski D (2017) Risk perception and disaster preparedness in immigrants and Canadian-born adults: Analysis of a national survey on similarities and differences. Risk Anal 37(12):2321–2333

Yong AG, Lemyre L (2019) Getting canadians prepared for natural disasters: a multi-method analysis of risk perception, behaviors, and the social environment. Nat Hazards 98(1):319–341

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

Paul Ralph PhD (British Columbia), is a professor of software engineering in the Faculty of Computer Science at Dalhousie University where his research centers on empirical software engineering, human-computer interaction and project management. Paul also co-chairs the ACM SIGSOFT Paper and Peer Review Quality Initiative. For more information please visit: https://paulralph.name.

Sebastian Baltes PhD (University of Trier), is a lecturer in the School of Computer Science at the University of Adelaide, Australia. His research empirically analyzes software developers' work habits to derive tool requirements and to identify potential process improvements. For more information please visit: https://empirical-software. engineering.

Gianisa Adisaputri Master of Emergency Management (Auckland University of Technology), MD (Islamic State University Syarif Hidayatullah Jakarta), is an emergency and disaster management consultant in Halifax, Canada. Her research interests include community resilience, disaster preparedness and emergency risk communication.

Richard Torkar PhD Blekinge Institute of Technology, is a professor of software engineering with the Software Engineering Division, Chalmers and University of Gothenburg; head of Software Engineering Division with the Department of Computer Science and Engineering; and senator with the Faculty Senate at the Chalmers University of Technology.


Vladimir Kovalenko PhD Candidate (Delft University of Technology), is a senior researcher at JetBrains Research in Amsterdam, The Netherlands, where he works on making software development process more efficient, in particular, by designing and building data-driven features for next-generation team collaboration tools. His research is dedicated to less studied aspects of design and implementation of data-driven software engineering tools.

Marcos Kalinowski PhD (COPPE/UFRJ), is a professor of software engineering and graduate program coordinator in the Department of Informatics at Pontifical Catholic University of Rio de Janeiro. His research interests include empirical methods in software engineering, software quality, and software engineering for artificial intelligence and digital transformation. For more information please visit: http://www.inf.puc-rio.br/~kalinowski.

Nicole Novielli PhD (University of Bari), is an assistant professor at the University of Bari. Her research interests intersect software engineering and affective computing, specifically focusing on emotion mining from software repositories and natural language processing of developers' communication traces, and using biometrics to classify developers' emotions during programming tasks.
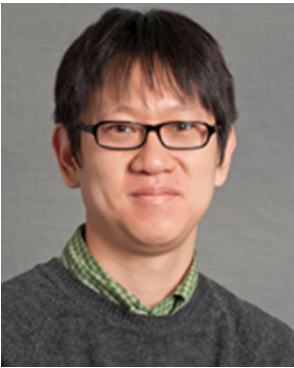
Shin Yoo PhD & MSc (King's College London), BSc (Seoul National University), is an associate professor at the School of Computing, KAIST, where he focuses on search-based software engineering research. Shin was Program Co-chair of IEEE ICST 2018, and is an editorial board member of TOSEM and EMSE. For more information, please visit: https://coinse.kaist.ac.kr/members/shin.yoo/.


Xavier Devroey Ph.D. (University of Namur), is a post-doctoral researcher in the software engineering research group (SERG) at Delft University of Technology. His research interests are search-based and model-based software testing, test suite augmentation, and variability-intensive systems testing. During the past three years, Xavier was involved in the EU Software Testing AMPlification (STAMP) project. For more information please visit: http://xdevroey.be.

Xin Tan PhD Candidate (Peking University), has research interests including software repository mining, empirical software engineering, and open source ecosystems. For more information please visit: https://sites.google.com/view/xintan.

Minghui Zhou PhD, is a professor in computer science at Peking University. Her main interest is software digital sociology. For more information please visit: http://sei.pku.edu.cn/~zhmh/.
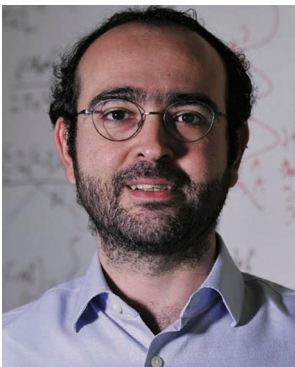

Burak Turhan PhD (Bogazici), is an associate professor in the Department of Software Systems & Cybersecurity at Monash University, and an adjunct professor at the University of Oulu. His research focuses on empirical software engineering, software analytics, quality assurance and testing, and human factors. For more information please visit: https://turhanb.net.

Rashina Hoda PhD (Victoria University of Wellington), B.Sc. Hons (Louisiana State University), is an Associate Dean (Academic Development) and an Associate Professor in software engineering at the Faculty of Information Technology at Monash University where her research focuses on human-centred software engineering, agile software development, and grounded theory. Rashina serves on the IEEE TSE reviewer board, the IEEE Software advisory panel, and as associate editor for JSS and on the organising committees for ICSE2021, XP2020, and ASE2020. For more information please visit: www.rashina.com.
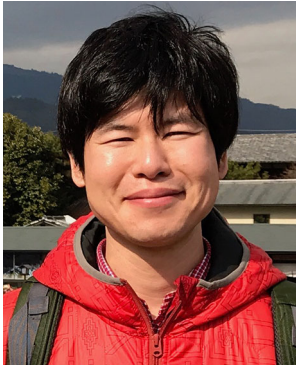
Hideaki Hata PhD (Osaka University), is an assistant professor in the division of information science at Nara Institute of Science and Technology, where his research centers on empirical software engineering, software ecosystems, human capital in software engineering, and software economics. He is an associate editor for IEICE Transactions on Information and Systems and has served on the PC of several conferences like ASE, MSR, and ICGSE. For more information please visit: https://hideakihata.github.io/.

Gregorio Robles PhD, is an associate professor at the Universidad Rey Juan Carlos, Madrid, Spain. Gregorio is specialized in free/open source software research. He is one of the founders of Bitergia, a software development analytics company. His homepage can be found at http://gsyc.urjc.es/~grex. Gregorio acknowledges the support of the Government of Spain through project "BugBirth" (RTI2018-101963-B-100).

Amin Milani Fard PhD (University of British Columbia), M.Sc. (Simon Fraser University), is an assistant professor in computer science at New York Tech Vancouver, and a visiting scholar at Simon Fraser University. His research and industry experience are in software engineering, data analysis, and cybersecurity. For more information please visit: http://www.ece.ubc.ca/~aminmf/.

Rana Alkadhi PhD (Technical University of Munich), is an assistant professor in computer science at King Saud University where her research centers on empirical software engineering, human aspects of software engineering and natural language processing. Rana has several publications in highly recognized outlets. For more information please visit: https://fac.ksu.edu.sa/ralkadi

Affiliations

Paul Ralph[1] · **Sebastian Baltes[2]** · **Gianisa Adisaputri[1]** · **Richard Torkar[3,4]** · **Vladimir Kovalenko[5]** · **Marcos Kalinowski[6]** · **Nicole Novielli[7]** · **Shin Yoo[8]** · **Xavier Devroey[9]** · **Xin Tan[10]** · **Minghui Zhou[10]** · **Burak Turhan[11,12]** · **Rashina Hoda[11]** · **Hideaki Hata[13]** · **Gregorio Robles[14]** · **Amin Milani Fard[15]** · **Rana Alkadhi[16]**

Gianisa Adisaputri
gianisaa@gmail.com

Richard Torkar
richard.torkar@cse.gu.se

Vladimir Kovalenko
vladimir.kovalenko@jetbrains.com

Marcos Kalinowski
kalinowski@inf.puc-rio.br

Nicole Novielli
nicole.novielli@uniba.it

Shin Yoo
shin.yoo@kaist.ac.kr

Xavier Devroey
X.D.M.Devroey@tudelft.nl

Xin Tan
tanxin16@pku.edu.cn

Minghui Zhou
zhmh@pku.edu.cn

Rashina Hoda
rashina.hoda@monash.edu

Hideaki Hata
hata@is.naist.jp

Gregorio Robles
grex@gsyc.urjc.es

Amin Milani Fard
amilanif@nyit.edu

Rana Alkadhi
ralkadi@ksu.edu.sa

1   Dalhousie University, Halifax, NS B3H 4R2, Canada
2   The University of Adelaide, Adelaide SA 5005, Australia
3   Chalmers and University of Gothenburg, Gothenburg, Sweden
4   Stellenbosch Institute for Advanced Study, Stellenbosch, South Africa
5   JetBrains, Amsterdam, The Netherlands
6   Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil
7   University of Bari Aldo Moro, Bari, Italy
8   KAIST, Daejeon, South Korea
9   Delft University of Technology, Delft, Netherlands
10 Peking University, Beijing, China
11 Monash University, Melbourne, Australia
12 University of Oulu, Oulu, Finland
13 Nara Institute of Science and Technology, Ikoma, Japan
14 Universidad Rey Juan Carlos, Móstoles, Spain
15 New York Institute of Technology, Vancouver, Canada
16 King Saud University, Riyadh, Saudi Arabia