Table 1: Inclusion and exclusion criteria

	Inclusion	Exclusion
Population	Post-secondary student populations	Secondary school students (and see "settings" below)
Exposure	Any COVID-19 mitigation implemented by the government or educational institution. (country or institutional level)	Studies without a COVID-19 mitigation. Studies which only referenced the "pandemic" broadly.
Comparator	At least two data collection measurements: either pre or post **and** during the COVID-19 mitigation. Also includes cross-sectional studies with a baseline mental health measurement e.g. pre-existing conditions.	Only one data collection measurement.
Outcome	Mental health outcomes, measured using validated or non-validated measurement tools.	Mental well-being outcomes, measured using validated or non-validated measurement tools.
Setting	Studies with college, undergraduate, and postgraduate students in post-secondary educational institutions as participants.	Studies with participants in residency training programs, co-op placements, high-school pupils who attend college to attain additional qualifications not available at their secondary school, education, students on work placement, internship, apprenticeships, and graduate schemes.
Study design	Studies in any design besides those specified in the exclusion criteria.	Modelling studies, opinion, editorials, reviews, and conference abstracts.
Geographical location	Studies conducted in any country or countries	No restrictions based on geographical location
Language	Studies published in English language	Studies published in language other than English

Table 2: Description of studies

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
Arad 2021	Longitudinal (pre/post) without control group. Also compares two groups at different time points (COVID and non-COVID)	Mitigation	Longitudinal - autumn 2019 (pre), spring 2020 (during); cross-sectional autumn and spring annually, 2016 - 2019 (pre-COVID)	44 (79.5% female, mean age [SD] 21.57 [1.90]) pre, 55 (89% female, mean age [SD] 22.62 [2.36]) post. First year students assessed as being socially anxious.	published	Israel	Lockdown (restriction to home), online learning and social distancing	Social anxiety
Baceviciene 2021	Longitudinal (pre/post) without control group	Mitigation	October 2019 (pre), February 2021 (during)	232 (79% female). Mean age 23.9 +/- 5.4 (19-39)	published	Lithuania	Lockdown	Disordered eating, binge drinking
Bolatov 2020	Cross-sectional - repeat survey conducted pre- and during COVID mitigations (overlap between samples unclear)	Mitigation	October/November 2019 (pre), April 2020 (during)	619 (75.0% female, mean age 19.14, range 17–27) - pre. 798 (75.7% female, mean age 20.31, range 17–33) - post. Medical students, years 1 to 5.	published	Kazakhstan	Online learning	Depression, anxiety, burnout, somatic symptoms
Bollen 2021	Cross-sectional - compares current and recalled outcomes	Mitigation	April/May 2020	1951	published	Belgium	Lockdown (mandated stay at home, ban on non-essential movement outside the home, campus closure)	Alcohol consumption
Conceicao-2021	Longitudinal (pre/post) without control group	COVID-19 pandemic	January, May and October 2019 (pre), June 2020 (during)	341 (75.4 % female, mean age 19.91 (SD 1.58))	preprint	Portugal	Lockdown and online learning	Depression, anxiety
Conrad 2021	Cross-sectional - compares outcomes between 2 groups	Mitigation	May-July 2020	791 (93% female) aged 18 - 30 years	published	USA	Mandated relocation from campus	Depression, anxiety, PTSD

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
Elmer 2020	Longitudinal (pre/post) without control group	COVID-19 pandemic	September 2019 (pre), April 2020 (during)	282 (21% female) pre. 212 post (% female not provided). Engineering and natural sciences students.	published	Switzerland	National lockdown (restricted social gatherings, closure of non-essential retail), campus closure, online teaching	Depressive symptoms, anxiety, perceived stress
Evans 2021	Longitudinal (pre/post) without control group	Mitigation	Autumn 2019 (pre), April/May 2020 (during)	254 (86% female. Mean (SD) age 19.76 (1.28)). First and second year psychology students.	published	UK	National lockdown (mandated stay at home apart from outside exercise once a day, all social gatherings prohibited, closure of non-essential businesses), online teaching	Alcohol use, depression, anxiety
Fruehwirth 2021	Longitudinal (pre/post) without control group	COVID-19 pandemic	October 2019/Feb 2020 (pre), June/July 2020 (during)	419 (70.4% female, ages 18-20 - mean (SE) 18.909 (0.019)). First year students.	published	USA	Campus closure (students sent home), online learning	Depression, anxiety
Hamza 2021	Longitudinal (pre/post) without control group	COVID-19 pandemic	May 2019 (pre), May 2020 (during)	733 (74 % female, mean age 18.52)	published	Canada	Lockdown (closure of non-essential businesses and public spaces), campus closure, online teaching	Perceived stress, non-suicidal self-injury, depressive symptoms, anxiety symptoms, borderline personality disorder symptoms, alcohol dependence, post-

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
								traumatic stress disorder
Horita et al., 2021	Cross-sectional with control group - compares two different groups at different time points (COVID and non-COVID)	Mitigation	April/May 2019 (pre), April/May 2020 (during).	766 (45.3% female) pre; 400 (56.5% female) post. First year students	published	Japan	Stay at home, online learning and campus closure	Depression, anxiety, eating concerns, alcohol use
Ji 2020	Longitudinal (high/moderate/low levels of lockdown) Three time points/three levels of quarantine without control group	COVID-19 pandemic	February 2020 (high quarantine, students on leave), March 2020 (moderate quarantine, students studying online from home), April 2020 (low quarantine, low COVID activity, students studying online from home)	4006 completed all three surveys. % female in each of the three surveys - 65.4, 64.7, 54.7. Mean (SD) age in each of the three surveys 21.3 (2.5), 21.2 (2.3), 20.9 (2.0). Majority were medical/clinical students.	published	China	High, moderate and low levels of quarantine (all levels involved students staying at home), online learning	Anxiety, OCD
Lawrence et al 2021	Cross-sectional - compares current and recalled outcomes	Mitigation		88 (93% female), mean age 28.97; SD 10.09. Social work students.	published	USA	Campus closure (including closure of student housing), stay at home order and online learning	Anxiety, depression, alcohol consumption
Li H.Y. 2020	Longitudinal (pre/post) without control group	Mitigation	December 2019 (pre), February 2020 (during)	555 (77% female. Mean age 19.6 yrs). First and second year students	published	China	Lockdown (restriction to home)	Anxiety and depression
Li, W. W et al., 2020	Longitudinal (pre/intense/less intense levels of lockdown). Three	Mitigation	November 2019 (pre), February/March 2020 (intense	173 (78% female). Mean (SD) age 19.81 (0.98)	published	China	Intense lockdown (students living at home), less intense	Anxiety and depression

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
	time points without control group.		lockdown), May/June 2020 (less intense lockdown)				lockdown but students still living at home	
Mehus 2021	Longitudinal (pre/post) without control group	COVID-19 pandemic	August/December 2019 (pre), April 2020 (during)	727 (64% female, age 18 - 23). First year students.	published	USA	Online learning	Anxiety and depression
Rosset-2021	Cross-sectional with control group - compares two different groups at different time points (COVID and non-COVID)	COVID-19 pandemic	Summer 2019 (pre), July 2020 (during)	155 (53% female). Mean age 20.85 (SD 2.19, range 17-30). Music/performing arts students.	published	Germany	University was partially open - strict hygiene regulations, restricted time schedules, social distancing, mask wearing and disinfection rules, blended learning.	Anxiety and depression
Savitsky 2020	Longitudinal (pre/post) without control group	Mitigation	March 2020 (during), May 2020 (at the end of lockdown)	113 Mean (SD) age 25.1 (3.3). First to fourth year students.	published	Israel	National lockdown, stay at home order, closure of essential businesses, campus closure, online learning	Anxiety
Sazakli-2021	Cross-sectional with control group - compares two different groups at different time points (COVID and non-COVID)	Mitigation	2017 (pre), April/May 2020 (during)	1989 (67 % female). Mean age 22.0 (SD 3.4). 91.1% undergraduates.	published	Greece	National lockdown, stay at home order, closure of essential businesses, campus closure, online learning	Anxiety and depression
Stubbe et al 2021	Longitudinal (pre/post) without control group	Mitigation	September 2019/February 2020 (pre), March/May 2020 (during)	98 (75% female. Mean age 19.89 (SD2.0)). 1st - 3rd year performing arts students	published	Netherlands	Closure of non-essential businesses, public spaces and institutions. Campus closure, prohibition of all performing arts	Mental health complaints, stress

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
							activities (classes, rehearsals, performances)	
Thomas 2021	Cross-sectional with control group - compares two different groups at different time points (COVID and non-COVID)	Mitigation	Before and after the implementation of home learning and curfew measures but dates not stated	119 (92.4% female, mean age: 20.85 (SD= 2.98)). 34 in pre-COVID group, 85 in post-COVID group.	published	UAE	National curfew, home learning	Depression
Wang-2020	Longitudinal (pre/post) without control group	Mitigation	February 2020 (within COVID period but students on leave), March 2020 (online learning)	1,172 (61 % female, age 18-22)	published	China	National lockdown (quarantine, social distancing), campus closure, online learning	Anxiety
Wilson 2021	Longitudinal (pre/post) without control group. Also compares two groups at different time points (COVID and non-COVID)	COVID-19 pandemic	Longitudinal - January 2020 (pre), April 2020 (after shift to online learning); cross-sectional January and April annually, 2015 - 2019 (pre-COVID)	832 (63.7% female) pre. 187 (68.4% female) post.	published	USA	Online learning	Depressive symptoms, stress
Yang-2021	Longitudinal (pre/post) without control group	COVID-19 pandemic	December 2018, June 2019, December 2019 (pre), June 2020 (during)	195 (58.5% female). First year students	published	China	Lockdown, quarantine, social distancing, and home confinement	Depressive symptoms, smart phone addiction
Zis 2021	Longitudinal (pre/post) without control group	Mitigation	January 2020 (pre), May 2020 (during)	154 (69.5% female. Mean age 22.6 ± 4.1 years (range: 18 to 52 years). Medical students, years 1 - 6.	published	Cyprus	Online learning	Mental health, burnout

Author	Study design	Focus of study (impact of COVID-19 or COVID-19 mitigation measures)	Data collection timing	n students	Publication status	Country	Mitigation	Outcome(s)
Zulevic 2021	Longitudinal (pre/post) without control group	Mitigation	December 2019/January 2020 (pre), June 2020 (during)	160 (68.8% female) Medical students, years 1 - 6 (47% first year)	published	Croatia	Online learning	Burnout

Table 3: Quality appraisal of included studies – cross-sectional studies

Author-Year	Were the criteria for inclusion in the sample clearly defined?	Were the study subjects and the setting described in detail?	Was the exposure measured in a valid and reliable way?	Were objective, standard criteria used for measurement of the condition?	Were confounding factors identified?	Were strategies to deal with confounding factors stated?	Were the outcomes measured in a valid and reliable way?	Was appropriate statistical analysis used?	Scoring	Overall appraisal
	Yes, No, Unclear, NA	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	Scoring out of total number of questions	*Very Low, Low, Moderate, High*
Lawrence et al 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8 out of 8	High
Bollen 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8 out of 8	High
Conrad 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8 out of 8	High

Table 4: Quality appraisal of included studies – quasi-experimental studies

Author-Year	Is it clear in the study what is the 'cause' and what is the 'effect' (i.e. there is no confusion about which variable comes first)?	Were the participants included in any comparisons similar? Yes: if the two groups were similar in characteristics	Were the participants included in any comparisons receiving similar interventions, other than the mitigation of interest? Yes: if the conditions both groups were under were similar i.e. setting	Was there a control group?	Were there multiple measurements of the outcome both pre and post the intervention/exposure?	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?	Were the outcomes of participants included in any comparisons measured in the same way?	Were outcomes measured in a reliable way?	Was appropriate statistical analysis used?	Scoring	No. of Yes	Overall appraisal
	Yes, No, Unclear, NA	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Yes, No, Unclear, NA*	*Out of number of questions*		*Very Low, Low, Moderate, High*
Thomas 2021	Unclear	Unclear	No	No	No	No	Yes	Yes	Yes	3 out of 9	3	Low
Ji 2020	Yes	Yes	Yes	No	No	Partial	Yes	Yes	Yes	6 out of 9	6	
Wilson 2021	Yes	Yes	No	No	Partial	Yes	Yes	Yes	Yes	7 out of 9	6	check: multiple pre only
Stubbe et al 2021	Yes	Yes	Yes	No	Partial	Partial	No	Yes	Yes	5 out of 9	5	Moderate (biggest flaw is low numbers and no confounding factors like performance schedule and how this affects stress are addressed)
Baceviciene 2021	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	6 out of 9	6	Moderate (overrepresentation of women in sample, a year between measurements ?other confounding factors important that not addressed)
Savitsky 2020	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6 out of 9	6	
Arad 2021	Yes	Yes	Yes	No	Partial	Yes	Yes	Yes	Yes	7 out of 9	7	moderate(Multiple(2-different types) surveys done post mitigation but not pre

Author-Year	Is it clear in the study what is the 'cause' and what is the 'effect' (i.e. there is no confusion about which variable comes first)?	Were the participants included in any comparisons similar? Yes: if the two groups were similar in characteristics	Were the participants included in any comparisons receiving similar interventions, other than the mitigation of interest? Yes: if the conditions both groups were under were similar i.e. setting	Was there a control group?	Were there multiple measurements of the outcome both pre and post the intervention/exposure?	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?	Were the outcomes of participants included in any comparisons measured in the same way?	Were outcomes measured in a reliable way?	Was appropriate statistical analysis used?	Scoring	No. of Yes	Overall appraisal
						Partial						mitigation,reliability could be good due to repetition of stat analysis keeping in mind confounders,but sample size is small with selection bias
Li H.Y. 2020	Yes	Yes	Yes	No	No	Partial	Yes	Yes	Yes	6 out of 9	6	Moderate,no control,good sample size but limited,narrow population considered(>females)
Hamza 2021	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	7 out of 9	7	
Fruehwirth 2021	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	7 out of 9	7	
Mehus 2021	Yes	Yes	Yes	No	Partial	Partial	Yes	Yes	Yes	6 out of 9	6	Moderate-no control check: multiple pre only
Elmer 2020	Yes	Unclear	Unclear	No	Partial	Partial	Yes	Yes	Yes	4 out of 9	4	- moderate - minimal demographic characteristics
Evans 2021	Yes	Yes	Yes	No	No	Partial	Yes	Yes	Yes	6 out of 9	6	
Zulevic 2021	Yes	Yes	Yes	No	No	Partial	Yes	Yes	Yes	6 out of 9	6	
Zis 2021	Yes	Yes	Yes	No	No	N/A	Yes	Yes	Yes	6 out of 9	6	
Bolatov 2020	Yes	Yes	Yes	No	No	Unclear	Yes	Yes	Yes	6 out of 9	6	
Horita et al., 2021	Yes	Unclear	Unclear	No	No	No	Yes	Yes	Yes	4 out of 9	4	low, no baseline characteristics, analysis not fully described,

Author-Year	Is it clear in the study what is the 'cause' and what is the 'effect' (i.e. there is no confusion about which variable comes first)?	Were the participants included in any comparisons similar? Yes: if the two groups were similar in characteristics	Were the participants included in any comparisons receiving similar interventions, other than the mitigation of interest? Yes: if the conditions both groups were under were similar i.e. setting	Was there a control group?	Were there multiple measurements of the outcome both pre and post the intervention/exposure?	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?	Were the outcomes of participants included in any comparisons measured in the same way?	Were outcomes measured in a reliable way?	Was appropriate statistical analysis used?	Scoring	No. of Yes	Overall appraisal
												publication less than two pages
Conceicao-2021	Yes	Yes	Yes	No	Partial	N/A	Yes	Yes	Yes	6 out of 9	6	
Yang-2021	Yes	Yes	Yes	No	Partial	N/A	Yes	Yes	Yes	6 out of 9	6	(preprint)
Wang-2020	Yes	Yes	No	No	No	N/A	Yes	Yes	Yes	5 out of 9	5	
Rosset-2021	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6 out of 9	6	
Li, W. W et al., 2020	Yes	Yes	No	No	No	Partial	Yes	Yes	Yes	5 out of 9	5	
Sazakli-2021	Yes	Unclear	Unclear	No	No	No	Yes	Yes	Yes	4 out of 9	4	-low (baseline characteristics of the comparison group not reported)

Table 5: Study results – depression prevalence

	Description of comparison	Instrument	Non-COVID comparator	During COVID	p
Bolatov 2020	Prevalence of depressive symptoms in a cohort of students receiving traditional face-to-face teaching pre-COVID (Oct/Nov 2019), compared to receiving online teaching during COVID (April 2020).	Patient Health Questionnaire-9 (PHQ-9)	49.3	27.6	p<0.001
Conceicao 2021	Prevalence of depressive symptoms in a cohort of students pre-COVID (May 2019) compared with during COVID (June 2020)	PHQ-9 ≥ 15	22.6	37	p<0.001
Evans 2021	Prevalence of clinical depression in a cohort of students in autumn 2019 (before lockdown) compared with lockdown conditions (April/May 2020)	Hospital Anxiety and Depression Scale (HADS)	13.8	34.3	p<0.001
Fruehwirth 2021	Prevalence of moderate to severe depression in a cohort of students pre-COVID/before stay at home order (October 2019 to Feb 2020) compared with during COVID (June/July 2020)	PHQ-8	21.5	31.7	p<0.001
Mehus 2021	Prevalence of possible moderate to severe depressive symptoms in a cohort of students before COVID (Aug/Dec 2019) compared with during COVID and COVID mitigations - online classes (April 2020)	PHQ-9 (prevalence is for scores of ≥10)	19.3	27.8	p < 0.001
Sazakli 2021	Prevalence of definite/possible depression in a cohort of students subject to lockdown with a different group of students surveyed in 2017, not subject to lockdown.	Hospital Anxiety and Depression Scale (HADS) Score ≥ denotes a possible or definite case	36	51.2	
Thomas 2021	Prevalence of severe depression in a cohort of students before and after the imposition of university curfew and home learning.	Beck Depression Inventory II (BDI-II) Prevalence is for severe depression	15	36.5	

	Description of comparison	Instrument	Non-COVID comparator	During COVID	p
Yang 2021	Prevalence of probable depression in a cohort of students before (Dec 2018, June 2019, Dec 2019) and after (June 2020) the introduction of lockdown and quarantine.	Chinese version of the 20-item Center for Epidemiologic Studies Depression Scale (CES-D). Scores ≥16 classified as probable depression	41.5, 45.6, 48.2 at the three pre-pandemic surveys	69.2	

Table 6: Study results – depression scores

	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	During COVID, but after partial easing of mitigation	p
Elmer 2020	Mean depression scores in a groups of students surveyed in April 2020, just after lockdown and switch to online learning, compared to a group of students surveyed in April 2019	Centre for Epidemiologic Studies Depression Scale (German version)	13.07		13.77		p=0.578
Evans 2021	Mean depression scores in a cohort of students in autumn 2019 (before lockdown) compared with lockdown conditions (April/May 2020)	Hospital Anxiety and Depression Scale (HADS)	4.33 (3.26)		6.31 (3.74)		p<0.001
Hamza 2021	Mean depressive symptoms score in a cohort of students surveyed in May 2020 compared with their scores in May 2019.	Centre for Epidemiologic Studies Depression Scale—Revised (CESD-R)	17.62 (13.46)		18.44 (13.24)		
Horita et al., 2021	Mean depression scores in a cohort of 1st year students subject to campus closure and online learning (April/May 2020) with the previous cohort of 1st year students (surveyed April/May 2019), not subjected to these changes.	Counselling Centre Assessment of Psychological Symptoms-Japanese (CCAPS-Japanese)	0.89 (0.72)		0.71 (0.61)		p<0.001
Lawrence et al 2021	Mean depressive symptoms scores in a group of students surveyed during stay-at-home compared with pre-stay-at-home, by asking participants to recall pre-stay-at-home anxiety	PHQ-9	3.63		9.1		p<0.0001
Li, W. W et al., 2020	Mean depression scores in a cohort of students before COVID (Nov 2019) during intense lockdown, students living at home on leave (Feb/March) and after the easing of some restrictions, whilst students still studying at home (May/June 2020)	Standardised Chinese version of the short Depression Anxiety Stress Scale (C-DASS21)	6.38	5.23		5.64	

	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	During COVID, but after partial easing of mitigation	p
Mehus 2021	Mean depression scores in a cohort of students before COVID (Aug/Dec 2019) compared with during COVID and COVID mitigations - online classes (April 2020)	PHQ-9	5.70 (5.33, 6.07)		6.83 (6.43, 7.23)		p < 0.001
Thomas 2021	Mean depression scores in a cohort of students before and after the imposition of university curfew and home learning.	Beck Depression Inventory II (BDI-II)	15.35 (9.96)		19.16 (12.49)		
Wilson 2021	Mean depressive symptom scores in a group of students surveyed in late January 2020 and again in late April 2020 (before and after shift to online learning) and further compared to similar cohorts of students between 2015 and 2019 surveyed at the same points in the semester each year.	CESD-7 - mean (SE)	Males: 4.6, SE 0.6 (3.9, SE 0.6) at the start (end) a normal semester. Females: 6.0, SE 0.5 (5.9, SE 0.4) at the start (end) of a COVID semester		Males: 4.8, SE 0.5 (5.4, SE 0.5) at the start (end) a normal semester. Females: 5.9, SE 0.6 (8.3, SE 0.4) at the start (end) of a COVID semester		ns (males). Sig (females) but p-value not stated
Yang 2021	Mean depression scores in a cohort of students before (Dec 2018, June 2019, Dec 2019) and after the introduction of lockdown and quarantine (June 2020).	Chinese version of the 20-item Center for Epidemiologic Studies Depression Scale (CES-D). Scores ≥16 classified as probable depression	14.61, 15.31, 15.93 at the three pre-pandemic surveys		19.08		p<0.001 (differences between the 3 pre-pandemic scores were ns)

Table 7: Study results – anxiety prevalence

	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	During COVID, but after easing of mitigation	p
Bolatov 2020	Prevalence of moderate to severe anxiety in a cohort of students receiving traditional face-to-face teaching pre-COVID (Oct/Nov 2019), compared to receiving online teaching during COVID (April 2020).	GAD-7	42.3		15.5		p < 0.001
Conceicao 2021	Prevalence of moderate/severe anxiety in a cohort of students pre-COVID (May 2019) compared with during COVID (June 2020)	GAD-7 (prevalence is for >10)	47.8		64.5		p < 0.001
Fruehwirth 2021	Prevalence of moderate to severe anxiety in a cohort of students pre-COVID/before stay at home order (October 2019 to Feb 2020) compared with during COVID (June/July 2020)	GAD-7	18.1		25.3		0.072
Mehus 2021	Prevalence of possible anxiety disorders in a cohort of students before COVID (Aug/Dec 2019) compared with during COVID and COVID mitigations - online classes (April 2020)	GAD-7 (8 or more indicating possible anxiety disorders)	24.3		29.6		p=0.003
Savitsky 2020	Prevalence of moderate to severe anxiety in a cohort after the easing of COVID restrictions (May 2020) compared with during the height of COVID restrictions (March 2020)	GAD-7 NB comparison is during to post. Prev is scores of 10 or more			42.8	34.9	
Sazakli 2021	Prevalence of possible or definite anxiety in a group of students subject to lockdown with a different group of students surveyed in 2017, not subject to lockdown.	Hospital Anxiety and Depression Scale [HADS] (Greek version). For each sub-scale, score 0-7 = non-case, 8-10 = possible case, ≥11 case.	56.7		35.8		

	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	During COVID, but after easing of mitigation	p
Wang 2020	Prevalence of anxiety in a cohort of students in mid-February 2020 (within the COVID period but students were on leave from university) compared to mid-March 2020 (after the start of the new semester, with online learning).	SAS Zung Self-Rating Anxiety Scale (50 - 59 = mild anxiety, 60 - 69 = moderate anxiety, ≥70 = severe anxiety)		15.7	18.86		p<0.05

Table 8: Study results – anxiety scores

Study	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	Post-COVID or post-COVID mitigation	p
Conceicao 2021	Mean anxiety score in a cohort of students pre-COVID (May 2019) compared with during COVID (June 2020)	GAD-7	9.98 (5.9)		12.15 (6.5)		p < 0.001
Elmer 2020	Mean anxiety scores in a group of students surveyed in April 2020, just after lockdown and switch to online learning, compared to a group of students surveyed in April 2019	GAD-7 (German version)	4.41		4.11		p=0.606
Evans 2021	Mean anxiety scores in a cohort of students in autumn 2019 (before lockdown) compared with lockdown conditions (April/May 2020)	HADS >=8 for generalised anxiety disorder	9.35 (4.28)		9.42 (4.47)		0.782
Fruehwirth 2021	Mean anxiety scores in a cohort of students pre-COVID/before stay at home order (October 2019 to Feb 2020) compared with during COVID (June/July 2020)	GAD-7	5.19 (SE0.155)		5.413 (0.224)		ns
Hamza 2021	Mean anxiety score in a cohort of students surveyed in May 2019 compared with their scores in May 2019.	GAD-7	6.68 (5.53)		6.39 (5.46)		
Horita et al., 2021	Mean anxiety scores in a cohort of 1st year students subject to campus closure and online learning (April/May 2020) with the previous cohort of 1st year students (surveyed April/May 2019), not subjected to these changes.	Counseling Center Assessment of Psychological Symptoms—Japanese	1.02 (0.63)		0.92 (0.62)		p<0.01
Horita et al., 2021	Mean social anxiety scores in a cohort of 1st year students subject to campus closure and online learning (April/May 2020) with the previous cohort of 1st year students (surveyed April/May 2019), not subjected to these changes.	Counseling Center Assessment of Psychological Symptoms—Japanese (social anxiety sub-scale)	2.01 (0.89)		1.86 (0.88)		p<0.05

Study	Description of comparison	Instrument	Pre-COVID	During COVID, pre-mitigation	During COVID, during mitigation	Post-COVID or post-COVID mitigation	p
Ji 2020	Mean anxiety scores in a cohort of students surveyed during winter break under high level quarantine (Feb 2020 – during COVID, pre-mitigation), whilst taking online courses at home under a moderate level of quarantine (March 2020 – during COVID, during mitigation) and whilst taking online courses at home under a low level of quarantine, during a period when no new cases were occurring (April 2020 – post-COVID or post-COVID mitigation)	SAS Zung Self-Rating Anxiety Scale		36.3 (7.4)	35.2 (7.6)	35.6 (7.9)	<0.001
Lawrence et al 2021	Mean anxiety scores in a group of students surveyed during stay-at-home compared with pre-stay-at-home, by asking participants to recall pre-stay-at-home anxiety	GAD-7	4.03 (4.36)		9.11 (6.07)		<0.0001
Li, W. W et al., 2020	Mean anxiety scores in a cohort of students before COVID (Nov 2019) during intense lockdown, students living at home on leave (Feb/March) and after the easing of some restrictions, whilst students still studying at home (May/June 2020)	Anxiety - Chinese version of the short Depression Anxiety Stress Scale (C-DASS21)	9.24		5.12	6.69	
Mehus 2021	Mean anxiety scores in a cohort of students before COVID (Aug/Dec 2019) compared with during COVID and COVID mitigations - online classes (April 2020)	GAD-7	5.07 (4.73, 5.41)		5.67 (5.30, 6.04)		p < 0.001
Savitsky 2020	Mean anxiety scores in a cohort after the easing of COVID restrictions (May 2020) compared with during the height of COVID restrictions (March 2020)	GAD-7 NB comparison is during to post. Prev is scores of 10 or more			9	7.5	<0.0001
Wang 2020	Mean anxiety score in a cohort of students in mid-February 2020 (within the COVID period but students were on leave from university) compared to mid-March 2020 (after the start of the new semester, with online learning).	SAS Zung Self-Rating Anxiety Scale		40.39 (9.98)	40.77 (10.51)		p < 0.05

Table 9: Study results - other mental health outcomes

Mental disorder	Measure	Study	Finding	
Anxiety and depression combined	Anxiety and depression, measured by PHQ-4	Li H.Y.	Mean (SD) anxiety and depression scores in a cohort of 1st and 2nd year students were 0.95 (0.65) pre-COVID in Dec 2019, compared with 0.76 (0.61) in February 2020, during home confinement (p<0,001).	Improvement
	Combined depression and anxiety, measured by 8-item symptom checklist (SCL-8) derived from SCL-25	Rosset	Mean (SD) anxiety and depression scores were 2.68 (0.89) in a pre-COVID group of students, compared with 2.75 (0.92) in a different but comparable group surveyed during the pandemic. Differences were not significant T(153) = −0.48, n.s.).	No change
Somatic symptoms	Patient Health Questionnaire-15 (PHQ-15)	Bolatov	Prevalence of somatic symptoms was 63.6% during face-to-face learning, falling to 19.4% during online learning (p<0.001)	Improvement
Mental health complaints	Oslo sports trauma research centre questionnaire on health problems (OSTRC)	Stubbe	Prevalence was 21.4% (Sep to Nov 2019), 24.5% (Dec 2019 to Feb 2020) and 27.6% (March - May 2020). Differences between the first and third time periods and between the second and third time periods were both significant (chi square 11.73, p<0.001 and chi square 19.45, p<0.001 respectively)	Deterioration
Mental health	Short Form Health Survey (SF-36)	Zis	Mental health significantly deteriorated in the overall sample (scores falling from 58.8 +/- 21.6 to 48.3 +/- 23, p<0.001). Mental health deteriorated in all age groups, reaching statistical significance in years 1 (p<0.001), 3 (p=0.008) and 6 (p=0.001)	Deterioration

Mental disorder	Measure	Study	Finding	
Post-traumatic stress disorder (PTSD)	PTSD Checklist (PCL)	Hamza	Students with preexisting mental health concerns showed significant improvement during the pandemic (mean (SD) 2.22 (0.94) compared with one year prior (2.35 (0.95)). In contrast, students without preexisting mental health concerns showed significant worsening during the pandemic (mean (SD) 1.53 (0.6) compared with one year prior (1.42 (0.5)), which coincided with increased social isolation among these students.	Improvement in those with pre-existing mental health conditions; deterioration in those without pre-existing mental health conditions
	PTSD Checklist—Civilian Version (PCL-C)	Conrad	No statistically significant association between being required to relocate and PTSD	No change
Non-suicidal self-injury (NSSI)	Inventory of Statements about Self-Injury (ISAS)	Hamza	No significant change in either group (i.e. those with and without pre-existing mental health conditions)	No change
Borderline personality disorder symptoms (BPD)	McLean Screening Instrument for borderline personality disorder (MSI-BPD)	Hamza	No significant difference in students with preexisting mental health concerns (mean (SD) 3.74 (2.64) compared with one year prior (4.02 (2.67)). In contrast, students without preexisting mental health concerns showed significant worsening during the pandemic (mean (SD) 1.63 (1.94) compared with one year prior (1.24 (1.55)), which coincided with increased social isolation among these students.	No change in those with pre-existing mental health conditions; deterioration in those without pre-existing mental health conditions
Obsessive-Compulsive Disorder (OCD)	Yale-Brown Obsessive-Compulsive Scale (Y-BOCS). A score of ≥ 16 denotes possible OCD	Ji	OCD scores were significantly higher at survey 1 (11.3 %) (high quarantine) than at surveys 2 (3.6 %) and 3 (3.5 %). P < 0.001	More intense lockdown conditions associated with higher OCD scores

Mental disorder	Measure	Study	Finding	
Disordered eating	Eating Disorder Examination Questionnair 6.0 EDE-Q6	Baceviciene 2021	No significant change in pre/post mean scores in either males (p=0.074) or females (p=0.594)	No change

Table 10: Study results - wellbeing outcomes

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
General health and wellbeing	Baceviciene 2021	Self-reported health	Single question, four-point response: "How would you describe your general health during the last 12 months?"	Deteriorated in men (p=0.008)	no change (women) worse (men)
	Baceviciene 2021	Overall Quality of Life	World Health Organization Quality of Life-BREF Questionnaire (WHOQOL-BREF). Lithuanian version	Overall quality of life (p=0.024) increased slightly but significantly in women. No change in men	better (women) no change (men)
Stress	Elmer 2020	Stress	Perceived Stress Scale (German version)	Between-cohort comparison: No significant difference between April 2019 cohort (14.76) and April 2020 cohort (15.12) p=0.757. Within-cohort comparison: increase in percieved stress between September and April (Mdiff = 2.67, SE = 0.40, t(208) = 6.64, p < .001, d = .23)	worse
	Hamza 2021	Recent stressful experiences	Inventory of College Students' Recent Life Experiences (ICSRLE) (assesses seven stressors relevant to university students)	Study found that several stressors decreased for students during the pandemic (e.g. having too many things to do, academic stressors, insufficient sleep).	better
	Hamza 2021	Stress	Perceived Stress Scale 10 (PSS-10)	Students with preexisting mental health concerns showed significant improvement during the pandemic (mean (SD) 3.25 (0.6) compared with one year prior (3.36 (0.58)). In contrast, students without preexisting mental health concerns showed significant worsening during the pandemic (mean (SD) 2.81 (0.63) compared with one year prior (2.70 (0.57)), which coincided with increased social isolation among these students.	better for students with pre-existing mental health conditions; worse for students without pre-existing mental health conditions
	Li, W. W et al., 2020	Stress	Depression Anxiety Stress Scale (C-DASS21) Chinese version	Stress scores peaked in pre-lockdown period (11.05), declining to lowest level (7.35) during intense lockdown, then increasing (8.44) post-lockdown.	V-shaped (worst pre-lockdown, best during "intense" lockdown, intermediate when lockdown relaxed)

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Stubbe et al 2021	Stress	Visual analogue scale was used to assess stress with scores ranging from 0 (no stress) to 100 (extreme amount of stress).	Mean stress scores were significantly lower for February (35.20) and March (36.41) when compared to the overall mean (40.38).	worse
	Wilson 2021	Stress	Percieved Stress Scale (PSS4)	Significant increase in perceived stress for both genders for the COVID period compared to the non-COVID period. Males: Mean PSS4 was 5.1, SE 0.3 (5.0, SE 0.4) at the start of a normal (COVID) semester and 5.2, SE 0.3 (6.8, SE 0.4) at the end of a normal (COVID) semester. These differences were significant. Females: Mean PSS4 was 6.2, SE 0.2 (6.4, SE 0.3) at the start of a normal (COVID) semester and 6.2, SE 0.2 (8.4, SE 0.3) at the end of a normal (COVID) semester.	worse
Sleep	Baceviciene 2021	Sleep duration	Single question regarding a participant's sleep duration in hours per day	Significantly longer in women (p<0.001)	better (women) no change (men)
	Evans 2021	Sleep quality	Pittsburgh Sleep Quality Index (PSQI)	There was no decline in sleep quality	no change
	Stubbe et al 2021	Sleep	Single-item numeric rating scale, where 1 is the worst possible sleep and 10 is the best	Sleep scores were significantly improved during the pandemic compared to pre-pandemic.	better

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
Loneliness and isolation	Conrad 2021	Loneliness	Adapted 3-item version of the UCLA Loneliness Scale Short Form	Students mandated to relocate reported more COVID-19-related grief, loneliness, and generalized anxiety symptoms compared to those who did not even after controlling for the severity level of local COVID-19 outbreaks. Students who had to leave behind valuable personal belongings reported more COVID-19-related worries, grief, and depressive, generalized anxiety, and PTSD symptoms.	worse in those mandated to relocate compared to those not mandated to relocate
	Elmer 2020	Social networks/social interaction	Bespoke method for this study	Anaylsis of within-person comparisons found that co-studying networks became sparser compared to pre-pandemic, with more students studying alone. Exploratory analyses suggest that isolation in social networks, lack of interaction and emotional support, and physical isolation were associated with negative mental health trajectories.	worse
	Elmer 2020	Loneliness	UCLA Loneliness Scale (German version)	Students became more lonely between September and April - (Mdiff = 0.13, SE = 0.02, t(208) = 5.59, p < .001, d = .07).	worse
	Evans 2021	Loneliness	De Jong Gierveld Loneliness Scale	No overall effects were observed on loneliness and respondents reported good adaptation to COVID restrictions.	no change
	Hamza 2021	Social mistreatment (isolation and loneliness)	Inventory of College Students' Recent Life Experiences (ICSRLE) social mistreatment scale	Study compared outcomes in participants with and without pre-existing mental health problems. Participants with preexisting mental health concerns reported no change in their mental health, or in social isolation and loneliness, whereas those without preexisting mental health problems reported deteriorating social isolation and loneliness and deteriorating mental health. The authors hypothesise that students who may be most adversely impacted by social distancing and most at risk of psychological distress are those who are unaccustomed to isolation and being alone. Increasing social isolation predicted greater mental health symptoms during the pandemic. The study	no change in those with pre-existing mental health concerns. Worse in those without pre-existing mental health concerns

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
				also found that several stressors decreased for students during the pandemic (e.g. having too many things to do, academic stressors, insufficient sleep).	
	Mehus 2021	COVID-related isolation	Single-item question with five-point scale: "How much has the COVID-19 situation made you feel isolated or alone?"	Regression models investigated associations between COVID-19-related variables and depression/anxiety outcomes. The largest effect sizes were for feelings of isolation; feeling extremely isolated (relative to not at all) was associated with depression symptom severity (proportional change[95% CI] = 2.43[1.87, 3.15]), anxiety symptom severity (2.02[1.50, 2.73]), greater odds of moderate depression symptom onset (OR[95% CI] = 14.83[3.00, 73.41]), and greater odds of moderate anxiety symptom onset (24.74[2.91, 210.00]).	
	Stubbe et al 2021	Loneliness	De Jong Gierveld and Van Tilburg 11-item loneliness scale	75.5% of the participants dealt with moderate to very severe loneliness during the pandemic, but unclear what pre-pandemic levels were.	unclear
Health behaviours - alcohol	Baceviciene 2021	Alcohol - binge drinking	Single question from the Health Behaviour among Lithuatian Adult Population survey	no significant change in pre/post mean scores in either males (p=0.71) or females (p=0.179)	no change

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Bollen 2021	Alcohol consumption	Increased, decreased or unchanged	More students reported reduced consumption during lockdown (68.2%) than increased (17.2%) or unchanged (14.6%; χ2 = 1067.484, p < .001) consumption. Mean (SD) number of units per week was 11.3 (13.5) pre-lockdown compared with 3.6 (5.75) post.The COVID-19-related lockdown was associated with reduced alcohol consumption in nearly 70% of this student sample. Pre-pandemic drinking motives constitute reliable predictors of consumption during lockdown, as (1) strong enhancement and social motives among heavy drinkers were associated with lower alcohol use, and (2) coping motives, as well as social motives in low drinkers, were related to higher consumption.	better
	Evans 2021	Alcohol use	AUDIT-C (the first three questions of the AUDIT questionnaire)	4.53 +/- 2.85 (pre-lockdown, compared with 3.96 +/- 2.58 during lockdown – there was a significant decrease in alcohol use during lockdown (p<.001)	better
	Hamza 2021	Alcohol dependence symptoms	Alcohol Use Disorders Identification Test (AUDIT)	No significant change in either group	no change
	Horita et al., 2021	Alcohol use	Counseling Center Assessment of Psychological Symptoms—Japanese (CCAPS-Japanese)	There was no significant difference in means scores for alcohol use. Mean (SD) alcohol use 0.04 (0.17) in 2020 vs 0.03 (0.17) - ns	no change

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Lawrence et al 2021	Alcohol consumption	Alcohol Use Disorders Identification Test (AUDIT-C). 0–12 point scale, with 0 reflecting no alcohol use.	There was no significant increase in alcohol scores - mean (SD) 5.13 (2.13) pre to 5.56 (2.82) post, p=0.06)	no change
Health behaviours – food/eating	Baceviciene 2021	Fast food consumption	Food frequency questionnaire containing 19 different groups of foods from the national survey of Health Behavior among Lithuanian Adult Population, 2014	Fast food consumption fell significantly in men (p=0.044)	no change (women) better (men)
	Baceviciene 2021	Unhealthy nutritional habits	Food frequency questionnaire containing 19 different groups of foods from the national survey of Health Behavior among Lithuanian Adult Population, 2014	Frequency of unhealthy nutritional habits decreased in both genders (p= 0.002 in men and p<0.001 in women)	better
	Baceviciene 2021	BMI	kg/m2	Increased significantly in women (p=0.009)	worse (women) no change (men)

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Horita et al., 2021	Eating concerns	Counseling Center Assessment of Psychological Symptoms—Japanese (CCAPS-Japanese)	There was no significant difference in mean scores for eating concerns. Mean (SD) eating concerns (0.95(0.71) in 2020 vs 0.92 (0.70) in 2019) - ns	no change
Health behaviours – physical activity	Baceviciene 2021	Physical activity	Leisure Time Exercise Questionnaire (LTEQ)	Leisure-time physical activity increased significantly in men (p=0.003).	no change (women) better (men)
	Lawrence et al 2021	Physical activity	Modified Godin Leisure Time Exercise Questionnaire	There was no significant change in levels of physical activity.	no change
	Wilson 2021	Physical activity	Global Physical Activity Questionnaire	Study found a significant decline in physical activity among college students; however there was no significant interaction between COVID and the change in physical activity levels in relation to stress and depression (i.e. there was no significant association between the change in PA and mental health).	worse

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
Health behaviours – screen time	Baceviciene 2021	Screen time	Single question: "On average, during your typical day, how many hours do you spend browsing the internet on your smartphone or computer for non - academic purposes and communicating in social networks?"	Duration of time spent browsing the internet increased significantly in both genders (p<0.001).	worse
	Lawrence et al 2021	Screen time	Previously validated survey estimating hours per day spent on activities such as watching television, using a computer/tablet/smartphone, and playing video games (Utter et al., 2003. Journal of the American Dietetic Association, 103(10), 1298–1305)	Study found a significant increase in screen time (from 4.48 to 5.94 hours per day, p<0.0001).	worse
	Yang-2021	Smart phone addiction (SPA)	Mobile Phone Addiction Index Scale (17-item scale), higher score indicates a higher level of SPA	Smartphone addiction did not increase during the pandemic. The level of smartphone addiction at the initial pre-pandemic measurement was significantly higher than at any subsequent time (p<.05). There were no significant differences between smartphone addiction levels at subsequent pre-pandemic and post-pandemic measurements (p>.05). Boredom and emotional loneliness were positively associated with smartphone addiction during COVID-19. Quarantine and lockdown were not significantly associated with smartphone addiction.	no difference

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
Burnout and academic distress	Bolatov 2020	Burnout	Copenhagen Burnout Inventory (CBI-S)	prevalence of burnout syndrome was 27.6% during face-to-face learning, falling to 16.7% during online learning (p<0.001)	better
	Zis 2021	Burnout	Maslach Burnout Inventory—Student Survey (MBI-SS). There are three subscales - subscales: exhaustion (EX), cynicism (CY) and efficacy (EF).	Overall prevalence of burnout did not differ significantly between the two periods (pre-COVID-19 18.1% vs. COVID-19 18.2%) but there were some significant differences within specific year groups. Year 4 is the year when clinical training begins. Burnout among year 4 students was significantly lower during the COVID, compared to the pre-COVID period (40.7% during the pre-COVID-19 period compared with 16.7% during the COVID-19 period, p=0.011), reflecting the fact that during the COVID period, 4th year students did not start their clinical training and therefore were not having to deal with the associated stress. In contrast, burnout among year 6 students was significantly higher during the COVID, compared to the pre-COVID period (27.6% during the pre-COVID-19 period compared with 50% during the COVID-19 period, p=0.01), reflecting the fact that these students are just about to start working as junior doctors. Doctors, particularly at the first stages of their careers, suffer from increased stress due to the responsibilities they have to colleagues and to patients. Clinical experience often reduces such stressors. The fact that the clinical training of sixth-year doctors stopped and went virtual had a negative effect on their psychology and confidence.	no difference overall but better for year 4 and worse for year 6 students

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Zulevic 2021	Burnout	Oldenburg Burnout Inventory (validated in Croatian) - separate measures for disengagement and exhaustion, and modified Copenhagen Burnout Inventory - separate measures for personal, studies, professor, student and patient-related burnout.	No significant difference on any burnout measure between pre- and post-intervention questionnaires.	no difference
	Horita et al., 2021	Academic distress	Academic distress	Study found that levels of academic distress were significantly higher in 2020 than in 2019 (p<0.001). Authors hypothesise that this might be reflective of the need to adapt to an unfamiliar online learning environment.	worse
Body image	Baceviciene 2021	Appearance evaluation	Multidimensional Body–Self Relations Questionnaire–Appearance Scales (MBSRQ-AS). Lithuanian version.	Appearance evaluation (p=0.037) increased slightly but significantly in women	better (women) no change (men)
	Baceviciene 2021	Media pressure	Sociocultural Attitudes Towards Appearance Questionnaire-4 (SATAQ-4)	Significantly increased in women (p=0.031).	worse (women) no change (men)

Wellbeing category	Study	Outcome	Instrument	Results	Better, worse or no change
	Baceviciene 2021	Internalisation of thin/low body fat ideals	Sociocultural Attitudes Towards Appearance Questionnaire-4 (SATAQ-4)	Significantly increased during lockdown in both genders ($p < 0.001$).	worse

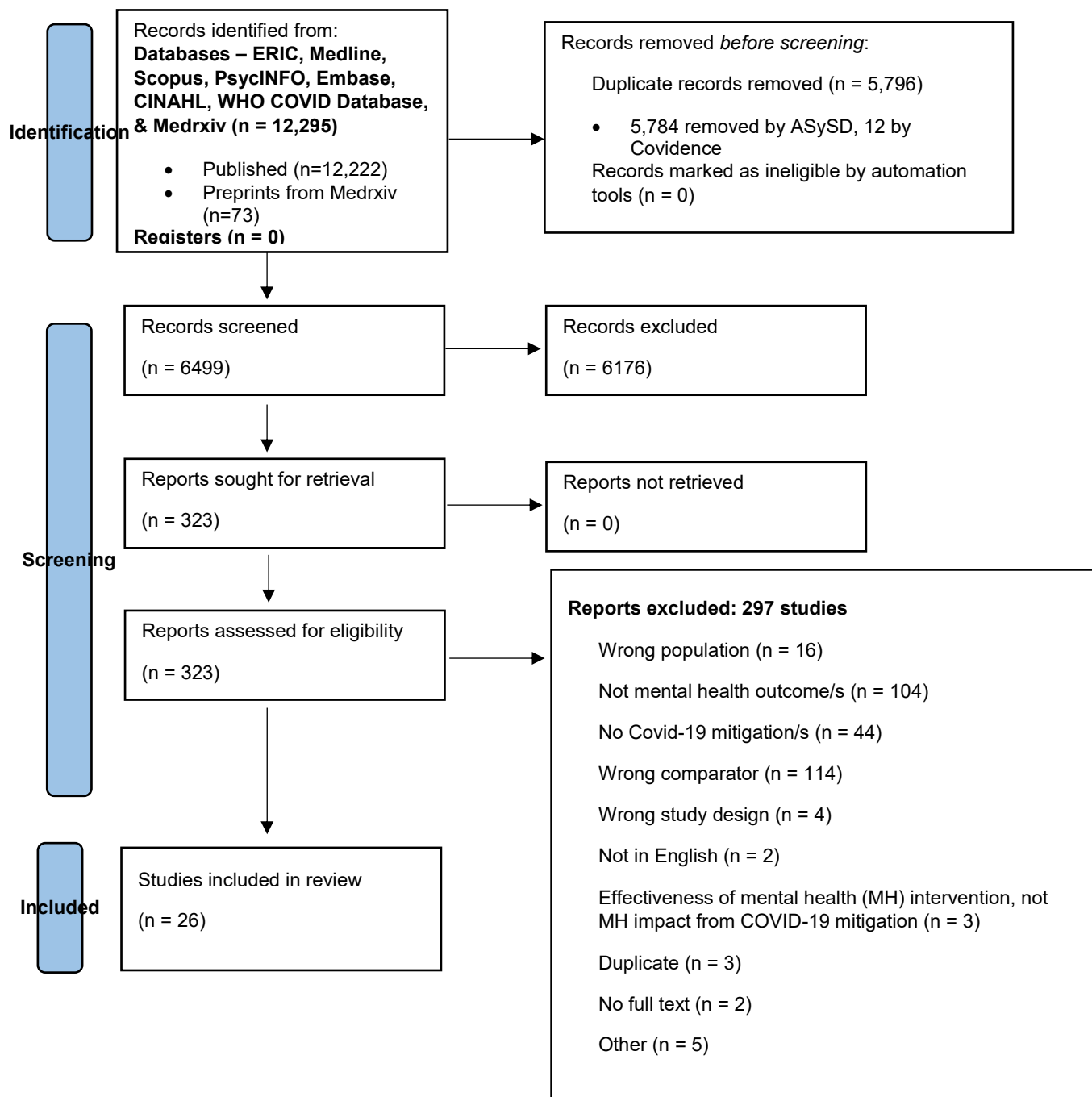**Figure 1: PRISMA diagram**

PRISMA Flow Chart

Figure 2: Visual summary of all mental health and wellbeing outcomes included in the review (panel A – full sample; panel B – sensitivity analysis with highest quality studies)

Panel A

Study	Anxiety	Depression	Anxiety and depression combined	General mental health	Borderline personality disorder (BPD) symptoms	Disordered eating	Non-suicidal self-injury (NSSI)	Obsessive-compulsive disorder (OCD)	Post-traumatic stress disorder (PTDD)	Stress	Recent stressful experiences	Loneliness	Social networks/social interaction	Sleep	Alcohol behaviours	BMI	Fast food consumption	Unhealthy eating	Eating concerns	Screen time	Physical activity	Burnout and academic distress	Body image - appearance evaluation	Body image - internalisation of thin ideals	Body image - media pressure	General health and wellbeing - overall quality of life	General health and wellbeing - self-reported health
Arad 2021	21																										
Baceviciene 2021														1		4	2				2			1	4	1	3
Bolatov 2020				20																							
Bollen																											
Conceicao-2021	8																										
Conrad 2021	9	9										9															
Elmer 2020	14	14																									
Evans 2021																											
Fruehwirth 2021	8											22															
Hamza 2021	5	5			6				5	5		6															
Horita et al., 2021	15																										
Ji 2020	16							16																			
Lawrence et al 2021																											
Li H.Y.																											
Li, W. W et al., 2020	7	7									17																
Mehus 2021	8	8										12, 22															
Rosset																											
Savitsky 2020	8, 10																										
Sazakli-2021	7	8																									
Stubbe				19								13															
Thomas 2021		7,9																									
Wang-2020	8																										
Wilson 2021																											
Yang-2021		8										22															
Zis				18																		11					
Zulevic 2021																											

Panel B

Study	Anxiety	Depression	Anxiety and depression combined	General mental health	Borderline personality disorder (BPD) symptoms	Disordered eating	Non-suicidal self-injury (NSSI)	Obsessive-compulsive disorder (OCD)	Post-traumatic stress disorder (PTDD)	Stress	Recent stressful experiences	Loneliness	Sleep	Alcohol behaviours	BMI	Fast food consumption	Unhealthy eating	Screen time	Physical activity	Burnout and academic distress	Body image - appearance evaluation	Body image - internalisation of thin ideals	Body image - media pressure	General health and wellbeing - overall quality of life	General health and wellbeing - self-reported health
Arad 2021	21																								
Baceviciene 2021						■							1	■	4	2	■	■	2		1	■	4	1	3
Bolatov 2020	■	■		20																■					
Conceicao-2021	8	■																							
Conrad 2021	9	9							■			9													
Evans 2021	■	■										■	■	■											
Fruehwirth 2021	8	■										22													
Hamza 2021	5	5			6		■		5	5	■	6		■											
Ji 2020	16							16																	
Li H.Y.			■																						
Mehus 2021	8	8										12, 22													
Rosset			■																						
Savitsky 2020	8, 10	■																							
Wilson 2021									■									■							
Yang-2021		8										22													
Zis				18																11					
Zulevic 2021																									

Legend:
- Significant improvement (COVID vs comparator)
- Potential improvement (COVID vs comparator) but p-values not provided
- Improvement in some sub-groups, no change in others (COVID vs comparator)
- Mixed outcomes (improvement in some, deterioration in others, COVID vs comparator)
- Deterioration in some sub-groups, no change in others (COVID vs comparator)
- Potential deterioration (COVID vs comparator) but p-values not provided
- Significant deterioration (COVID vs comparator)
- Inconclusive/no change

Notes:

[1] Improvement in women, no change in men

[2] Improvement in men, no change in women

[3] Deterioration in men, no change in women

[4] Deterioration in women, no change in men

[5]Improvement in students with pre-existing mental health conditions, deterioration in students without pre-existing mental health conditions

[6]No change in students with pre-existing mental health conditions, deterioration in students without pre-existing mental health conditions

[7]P-values not provided, so result is tentative

[8]Study reports both prevalence and mean score

[9]Comparison is between those mandated to relocate from campus (intervention) and those not required to relocate (control)

[10]Comparison is between lockdown (intervention) and post-lockdown (control)

[11]Burnout reduced among year 4 students who had clinical placements cancelled but increased among year 6 students about to graduate and start work as junior doctors

[12]Feelings of isolation predicted poor mental health outcomes, but unclear if feelings of isolation were linked to COVID mitigations

[13]High levels of loneliness during pandemic period, but levels of loneliness pre-pandemic not provided

[14]No change in cross-sectional comparison, deterioration in longitudinal comparison

[15]Significant improvement in both anxiety and social anxiety scores

[16]Study compares anxiety scores under different levels of quarantine. Outcomes were significantly worse during initial, intense lockdown when students were on winter break, compared with later, less intense lockdown, when they were studying online from home.

[17]V-shaped (worst pre-lockdown, best during "intense" lockdown, intermediate when lockdown relaxed) but p-values not provided

[18]Mental health

[19]Mental health complaints

[20]Somatic symptoms

[21]Socially anxious students became significantly less anxious from autumn to spring in a pre-COVID cohort but anxiety levels remained high in the COVID cohort

[22]Investigated as a risk factor