The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry 64:8 (2023), pp 1115–1139 doi:10.1111/jcpp.13754

Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms

Karima Susi,[1,2,3,4] Francesca Glover-Ford,[1,2] Anne Stewart,[2,5] Rebecca Knowles Bevis,[1] and Keith Hawton[2,6]

[1]Oxford Institute of Clinical Psychology Training and Research, Isis Education Centre, Warneford Hospital, Oxford Health NHS Foundation Trust and University of Oxford, Oxford, UK; [2]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK; [3]Child and Adolescent Mental Health Services, Leicestershire Partnership NHS Trust, Leicester, UK; [4]Step 4 Psychological Therapies, Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, UK; [5]Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK; [6]Centre for Suicide Research, Department of Psychiatry, University of Oxford, Oxford, UK

Background: Viewing self-harm and suicide-related images online can precede these behaviours. We reviewed studies of potential impacts and mechanisms associated with viewing self-harm-related images on the internet and social media. **Method:** CINAHL, Cochrane Library, EMBASE, HMIC, MEDLINE, PsycArticles, PsycINFO, PubMed, Scopus, Sociological Abstracts and Web of Science Core Collection databases were searched for relevant studies from inception to 22 January 2022. Inclusion criteria were English language, peer-reviewed, empirical studies with data related to impacts of viewing self-harm images or videos on the internet or social media. Quality and risk of bias were assessed using Critical Appraisal Skills Programme tools. A narrative synthesis approach was used. **Results:** Of the 15 identified studies, all found harmful effects of viewing self-harm-related images online. These included escalation of self-harm, reinforcement of engagement behaviours (e.g. commenting and sharing images), encouragement of social comparison (comparing own self-harm with others), development of a self-harm identity, social connection perpetuating or escalating self-harm, and emotional, cognitive, and physiological impacts triggering self-harm urges and acts. Nine studies found protective effects, including self-harm mitigation or reduction, promotion of self-harm recovery, encouraging social connection and help-giving, and emotional, cognitive and physiological impacts mitigating or reducing self-harm urges and acts. Causality of impact was not determined in any study. Most of the studies did not explicitly evaluate or discuss potential mechanisms. **Conclusions:** Viewing self-harm images online may have both harmful and protective effects, but harmful effects predominated in the studies. Clinically, it is important to assess individual's access to images relating to self-harm and suicide, and the associated impacts, alongside pre-existing vulnerabilities and contextual factors. Higher quality longitudinal research with less reliance on retrospective self-report is needed, as well as studies that test potential mechanisms. We have developed a conceptual model of the impact of viewing self-harm images online to inform future research. **Keywords:** Self-harm; self-injury; self-harm images; social media; internet usage.

Introduction

Self-harm and suicide are serious public health issues affecting young people, with suicide being the third leading cause of death among 15- to 19-year olds (World Health Organisation [WHO], 2019), and self-harm a major risk factor for suicide in children and adolescents (Hawton, Bale, et al., 2020). In this review, self-harm is defined as nonfatal intentional self-injury or self-poisoning, irrespective of suicidal intent or other motives (Hawton et al., 2003; National Institute for Health and Care Excellence [NICE], 2022).

Findings from several studies have raised concerns that self-harm and suicidal content viewed online or on social media, particularly 'graphic' images or videos, may be harmful; for example, by normalising self-harm and discouraging help-seeking (Baker & Lewis, 2013; Bell, 2014; Brennan et al., 2022; Daine et al., 2013; Jacob, Evans, & Scourfield, 2017; Lewis, Heath, Sornberger, & Arbuthnott, 2012; Lewis & Seko, 2016; Marchant et al., 2017; Marchant, Hawton, Burns, Stewart, & John, 2021). The potential for harm is particularly concerning because self-harm and suicide-related internet sites are commonly accessed by young people engaging in self-harm and are associated with both increased self-harm and suicidal intent (Daine et al., 2013; Mars et al., 2015; Padmanathan et al., 2018). Accessing suicide-related content online can be an antecedent to suicide in young people (Rodway et al., 2022).

The contagion or copycat effects of self-harm and suicide-related content online are important to consider, particularly in relation to geographical (temporal) suicide clusters, which can spread via the internet and on social media (Hawton, Hill, et al., 2020). Several potential mechanisms have

Conflict of interest statement: See Acknowledgements for full disclosures.

been proposed to account for contagious aspects of suicide clusters (Haw, Hawton, Niedzwiedz, & Platt, 2013; Hawton, Hill, et al., 2020), such as normalisation (descriptive norms; Cialdini, Reno, & Calgren, 1990; Rimal & Real, 2003) where suicidal behaviour appears to be more common in the community than it actually is, social learning (i.e. imitation; Bandura, 1977), and social positive reinforcement (Jarvi, Jackson, Swenson, & Crawford, 2013; Nock & Prinstein, 2004; Table S1). These mechanisms are also likely to be relevant when considering the impact of viewing self-harm-related content online, although there has been little discussion regarding mechanisms in this area.

Protective or positive influences associated with self-harm-related content online have also been reported, such as access to social support (Bell, 2014; Daine et al., 2013; Lewis & Seko, 2016; Marchant et al., 2017, 2021; Robinson et al., 2016). In the last few years, guidelines to help young people communicate safely about suicide on social media have been developed (Robinson et al., 2018).

The emotionally provocative and triggering nature of images (including mental imagery) compared to verbal content has been well documented (Holmes & Mathews, 2010; Kensinger & Schacter, 2006; Winkielman & Gogolushko, 2018) and warrants further investigation in relation to self-harm. Reviews to date have combined analysis of textual and visual content, studied only one platform (e.g. Instagram), or focused more on the content of self-harm images and the motivations for posting, rather than the impact of viewing self-harm images or the potential mechanisms underlying these effects. Assessing the quality of evidence relating to the impacts and associated mechanisms of viewing self-harm images is important, as this has clear implications for clinical practice and legislation regarding online safety.

In this systematic review we therefore aimed to answer the following questions: (a) What are the potentially harmful and protective impacts of viewing self-harm images on the internet and on social media? (b) What are the potential mechanisms which may account for harmful and protective effects?

Method

Search strategy and selection criteria

This review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2009) and was preregistered on PROSPERO (ID CRD42019137674). KS conducted an electronic search for articles published from inception to 22 January 2022 in the following databases: CINAHL, Cochrane Library, EMBASE, HMIC, MEDLINE, PsycArticles, PsycINFO, PubMed, Scopus, Sociological Abstracts and Web of Science Core Collection. Keyword searching of titles and abstracts, as well as medical subject headings (MeSH), where possible, were conducted, following librarian guidance, to maximise search sensitivity using:

1. Self-harm-related terms:
   (selfharm* OR 'Self Harm*' OR 'Self-Harm' OR 'Self Injur*' OR 'Self-Injur*' OR Cutting OR 'Non Suicidal Self Injur*' OR 'Non-Suicidal Self Injur*' OR 'Non-suicidal self-injur*' OR 'Self Destructive Behavio?r' OR 'Self-Destructive Behavio?r' OR 'Self Injurious Behavio?r' OR 'Self-Injurious Behavio?r' OR 'Self Inflicted Wound*' OR 'Self-Inflicted Wound*' OR 'Self Poisoning' OR Overdos* OR 'Self-Poisoning' OR 'Head banging' OR 'Suicidal Ideation' OR 'Attempted Suicide' OR Suicid* OR 'Suicidal Behavio?r*' OR Parasuicide OR 'Self Mutilation' OR 'Self-Mutilation' OR automutilation OR NSSI OR SIB)
   Example MeSH: Self-injurious behaviour.

2. AND image terms:
   (image* OR video* OR picture* OR graphic* OR 'graphic image*' OR photo*6 OR 'digital image*' OR 'digital media' OR 'digital video*' OR 'image* shar*' OR 'video shar*' OR 'photo* shar*' OR galler*)
   Example MeSH: Photograph.

3. AND social media terms:
   ('social media' OR internet OR online OR web* OR blog OR Facebook OR Instagram OR twitter OR Tumblr OR snapchat OR 'online social network*' OR 'online community' OR 'social interaction' OR 'electronic communication' OR 'communications media' OR chatroom OR forum* OR chat OR 'social network*' OR tweet* OR virtual* OR Whatsapp OR Pinterest OR vine OR vlog OR YouTube OR Flikr OR Reddit)
   Example MeSH: Social media

Supplementary searches were conducted using Google Scholar, and through forward and backward citation searching of related reviews and studies included in this review. Contact with field experts identified studies recently published (or in press) and clarified eligibility questions.

The inclusion criteria for studies (Table S2) were: (a) peer-reviewed; (b) empirical research; (c) English language; (d) included data related to the viewing of self-harm images (e.g. images/videos of self-injury/self-poisoning or related items); (e) self-harm images appeared online or on social media and (f) findings related to the potential impact of, or responses to, viewing self-harm images. Studies were excluded if they only examined verbal content, recovery-based initiatives, impact of fictional self-harm images (e.g. Netflix), or did not assess the potential impact, such as content analysis only or algorithm studies. Age of participants was not an exclusion criterion as many studies included anonymous users.

Identified studies were imported into Mendeley and duplicates were removed. KS screened titles and abstracts, and FGF independently screened 10% of the papers to confirm decisions regarding inclusion/exclusion made at this stage and there were no disagreements. Where eligibility could not be determined, full-text articles were independently reviewed by KS and FGF. At the full-text screening stage, a rationale for each decision was independently recorded by KS and FGF and disagreements were resolved by discussion between KS and FGF using the inclusion/exclusion criteria as a guide (Table S2), or, if agreement between KS and FGF could not be reached, through group consensus with RKB, AS and KH. Inter-rater agreement between KS and FGF was 'good' before discussion (κ = .71), and 'very good' after discussion (κ = .95; Landis & Koch, 1977; as cited in Laerd Statistics, 2015).

Data extraction and analysis

A data extraction sheet was developed to facilitate extraction of relevant information, such as demographics, harmful and protective impacts of images, and potential mechanisms. KS completed full data extraction for all papers, and FGF independently extracted data relating to the main findings/themes, harmful and protective impacts, potential mechanisms, and strengths and limitations (the key domains and those open to subjective judgement) from all papers.

The heterogeneity of the methodology and findings precluded meta-analysis or meta-synthesis. Therefore, a narrative synthesis was conducted to 'tell the 'story' of the qualitative and quantitative findings collectively (Popay et al., 2006). The findings across studies were synthesised and grouped according to the different areas of potential impact (harmful and protective), the potential underlying mechanisms, and the quality appraisal. Themes were generated by following a bottom-up, iterative process to ensure they were data-driven. On completion of data extraction, a framework of harmful and protective impacts and potential mechanisms was devised and agreed by all authors (Table 1). Evidence for harmful and protective impacts was recorded where applicable. Results from studies relating to impact could be recorded in multiple areas (e.g. evidence of impact on emotions and self-harm behaviour).

A mechanism was defined as a process that could explain the relationship between exposure and impact or that may have contributed to a change in affect, cognition, or behaviour (Holmes, Blackwell, Burnett Heyes, Renner, & Raes, 2016). Potential mechanisms relating to the harmful or protective impacts of online exposure to self-harm images were conceptualised using definitions from related papers in the field (see Table 2; Table S1; Haw et al., 2013; Hawton, Hill, et al., 2020; Jarvi et al., 2013). Any mechanism that was either explicitly or implicitly mentioned by the study authors, as well as those that were deemed to be potential mechanisms by the review authors, were identified and labelled. The results focus on the key mechanisms discussed either explicitly or implicitly by the study authors. Potential mechanisms were not mutually exclusive or exhaustive, and multiple potential mechanisms could account for different effects.

Quality and risk of bias was assessed using the Clinical Appraisal Skills Programme checklists (CASP, 2018a, 2018b). The cohort study checklist was used for quantitative studies (CASP, 2018a). This was judged to be the most appropriate tool following communication with CASP as it evaluates the effects of harm (e.g. assesses measurement of exposure, outcome and follow-up). The qualitative checklist was used for qualitative studies (CASP, 2018b). The appraisal tool was selected according to how the study authors themselves defined their research. Where both quantitative and qualitative data analysis were reported (i.e. Sternudd, 2012) both checklists were used, and a combined quality rating was given. Applying conventional cut-offs (Darlow et al., 2012; Marchant et al., 2021), articles were rated as low (<50%), medium (50–75%), and high (>75%) quality depending on their overall CASP scores (see Table S3). KS and FGF independently completed quality ratings for all papers and any disagreements were resolved by discussion to reach consensus. Weighted kappa (κ_w) calculations with linear weights (Cicchetti & Allison, 1971) showed excellent inter-rater reliability (κ_w = .91; Fleiss, Levin & Paik, 2003; as cited in Laerd Statistics, 2021). RKB and AS completed full data extraction and quality analysis on seven of the papers as an additional reliability check. Findings from higher quality papers were given more weight in the results and discussion.

Results

Summary of studies

Fifteen studies were included in the review (Figure 1), comprising seven quantitative, seven qualitative, and one mixed-methods study (Table 2; see Table S4 for a more detailed summary). Most studies were cross-

Table 1 Definitions of potentially harmful and protective areas of impact

Area of potential impact	Potentially harmful impacts	Potentially protective impacts
1. Self-harm behaviour	Triggering or increasing self-harm	Mitigating or reducing self-harm
2. Engagement behaviours (e.g. sharing, liking, or commenting on self-harm images)	Reinforcing the creation/sharing of self-harm images which may encourage self-harm behaviours	Reinforcing the creation/sharing of recovery narratives, such as including messages of hope or recovery (healing wounds or scars)
3. Social comparison	Comparing own self-harm with images, leading to an escalation of self-harm or impacting on mood	Comparing own self-harm with images discouraging self-harm
4. Self-harm identity	Self-harm images encouraging a self-harm identity	Self-harm images encouraging self-expression, reflection, and recovery potentially leading to reduced self-harm, and increased sense of self-efficacy and control over self-harm
5. Social connection	Sense of peer-support, belonging, connection, identification and validation which encourages a self-harm group identity and a 'crowd' mindset of self-harming Help-giving that may discourage help-seeking/recovery or encourages self-harm in other ways (e.g. sharing methods) and may increase emotional strain by distracting from own needs or increasing the sense of responsibility for others	Sense of peer-support, belonging, connection, identification and validation, reducing feelings of isolation which may curb urges to self-harm Help-giving which may encourage help-seeking/recovery, reduce isolation or discourage self-harm
6. Emotion	Emotional change increasing the likelihood of self-harm (e.g. emotional dysregulation, lower mood, stigma and lower aversion)	Emotional change reducing the likelihood of self-harm (e.g. vicarious relief, aversion or disgust)
7. Cognition (self-harm thoughts including urges and mental images)	Increase in self-harm cognitions or urges, or thinking about self-harm as an acceptable behaviour	Reduction in self-harm cognitions or urges
8. Physiological effects	Physiological change (e.g. 'rush') leading to an increased urge to self-harm	Physiological change (e.g. 'rush') leading to reduction in the need to self-harm

Table 2 Summary of studies

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
1. Arendt et al. (2019) USA Quantitative Longitudinal (two-wave panel design)	Whether exposure to self-harm images on Instagram increased the risk of self-harm and suicidality	Self-harm images on Instagram (e.g. cutting)	Questionnaire measuring self-harm, suicidal ideation and plans, suicide risk, hopelessness, reasons for living, emotional disturbance, perspective taking, and presumed copycat influence	*Harmful impacts on self-harm, emotion, and cognition.* Exposure to self-harm images on Instagram was associated with suicidal ideation, self-harm, and emotional disturbance after controlling for exposure to similar sources. Exposure predicted self-harm and suicidality 1 month later *Potential mechanism(s): Assortative relating, emotional dysregulation, and social learning.* Exposure leads to contagion or imitation and emotional dysregulation either because of the exposure itself or because individuals are at a higher risk to begin with
2. Baker and Lewis (2013) Canada Qualitative Cross-sectional	Individuals' reactions to online NSSI (nonsuicidal self-injury) photos	NSSI photos on large NSSI e-community gallery (e.g. scratching, severe cuts, burns and scars)	Analysed 89 posts left by 88 authors on a message board depicting their self-reported reactions to NSSI image gallery	*Harmful impacts on self-harm, engagement behaviours, social comparison, social connection, and cognition.* 51.7% endorsed NSSI images as unhelpful and themes were: (a) social responsibility, (b) images can normalise, glamorise and encourage competition relating to self-harm, and (c) could trigger NSSI *Potential mechanism(s): Descriptive norms and comparison and competitiveness.* Images may reinforce NSSI through narrative reinforcement (descriptive norms) and competition *Protective impacts on self-harm, social connection, emotion, cognition and physiology.* 44.9% endorsed NSSI photos as helpful and themes were: (a) images show how bad NSSI can get, (b) enable identification and support with each other, and (c) curb urges to self-injure *Potential mechanism(s): Aversion* to self-harm arising from viewing images of scars or severe self-injury, and emotional regulation (e.g. relief) through viewing images as proxy to self-harm
3. Brown et al. (2018) Germany Quantitative Cross-sectional	Whether social reinforcement predicted positive associations of NSSI pictures and	NSSI wounds (mostly cutting) on Instagram found using German NSSI hashtags over 4-weeks	2,826 NSSI images (from 1,154 accounts) and 6,651 associated comments posted by other users	*Harmful impacts on engagement behaviours and social connection.* Pictures with more severe wounds and multiple methods of NSSI generated more comments. 0.5% of comments ($n = 33$) complimented the wound/picture. 6.8% of comments ($n = 450$) were abusive

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
	comments being posted			*Potential mechanism(s): Social positive reinforcement* might play a role in posting of more severe NSSI images Protective impact on social connection. 23.5% (n = 1,562) of comments were empathic, 11.6% (n = 770) were warnings asking the user to stop NSSI, and 6.9% (n = 462) of comments were offering help *Potential mechanism(s): Social positive reinforcement* might play a role in encouraging others to post images of NSSI for social connection functions
4. Brown et al. (2020) Germany Qualitative Cross-sectional	Adolescents' motivations for sharing NSSI pictures on Instagram, their reactions to viewing NSSI pictures online, and how they perceive comments on their own NSSI pictures	NSSI images/ videos on Instagram (mostly cuts but some reference to scars)	Retrospective self-reported impact of NSSI images measured through 59 online semi-structured interviews conducted via Instagram messaging app (identified participants from Brown et al., 2018)	*Harmful impacts on self-harm, engagement behaviours, social comparison, social connection and emotion.* Participants (30.5%) reported images could trigger self-harm. Some (6.8%) reported following others' NSSI accounts before posting their own. More severe NSSI images reported to elicit more severe emotional reactions and attention (54.2%). Users most commonly identified with others' images (35.6%) and compared ones' own wounds with others. *Potential mechanism(s): Emotional dysregulation, social positive reinforcement and social learning.* Posting more severe images can be socially reinforced by more attention. Viewing images may lead to social contagion and imitation and trigger self-harm *Protective impacts on self-harm, identity, social connection and emotion.* Participants (11.9%) reported others' repulsive graphic NSSI images sometimes deterred self-harm. All participants reported having received positive reactions to their own NSSI images (being offered help, connection, empathy) or wanting to offer help when seeing others' NSSI images (25.4%). Images were used for documentation, recovery, or as a way of offering help to others *Potential mechanism(s): Disgust and social positive reinforcement.* Viewing images sometimes deterred participants from self-harm. Positive feedback and social support in response to images might have positive effects on recovery

(continues)

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
5. Cha et al. (2016)[a] Conducted in USA [105 countries (64.2% USA residents)] Quantitative Cross-sectional Experimental (only Study 1 was web-based)	Examined impact of self-harm images (cut skin) versus words (suicide, death) on mood, desire to self-harm and die, using Implicit Association Tests (IAT) on web-based responders	Images of self-harm (cutting) compared with words (death, suicide) presented on web-based IAT	Self-report questionnaires: Before and after each IAT, participants in each group were asked about their desire to self-injure, desire to die, and mood on a Likert scale. Self-injurious Thoughts and Behaviors interview assessed NSSI, suicidal ideation, suicide plan, and suicide attempt	*Harmful impact on emotions*. In study 1 there was a small but significant mood decline for the self-injury IAT (images), but not suicide or death word IATs, even after controlling for self-harm history, suggesting that self-harm images may impact on mood more than words/phrases, particularly in females *Potential Mechanism(s): Emotion dysregulation*. Images related to self-injury and suicide may be more emotionally provocative than words, supporting neuroscience literature
6. Chen et al. (2021) China Qualitative Cross-sectional	Explored what triggers and maintains nonfatal self-harm among Chinese adolescents and help-seeking barriers	Pictures of self-harm on chat groups	Retrospective self-reported impact via semi-structured interviews	*Harmful impact on cognition*. Peers sharing self-harm images online inspired ideas of self-harm *Potential mechanism(s): Social learning*. Adolescents learned about self-harm from seeing it in online chat groups
7. Fu et al. (2013) China Quantitative Cross-sectional	Investigated (a) how people respond to self-presentation of self-harm; (b) how posts propagate online and (c) implications for suicide prevention	One picture of wrist cutting posted on Sina Weibo (Chinese social media platform)	Analysed content of responses to one wrist cutting picture (1,997 comments and 3,974 reposts)	*Harmful impacts on engagement behaviours and emotions*. The original post was reposted 3,974 times and received 1,997 comments by 3,696 microbloggers within 3 h, yielding 5,971 microblog pieces of content. 23.4% of responses evidenced a negative attitude (i.e. cynical, or indifferent comments, cyberbullying), 19.5% were emotional presentations of shock and 20.4% were merely reposts *Potential mechanism(s): Emotional dysregulation and social positive reinforcement*. Disconfirmation comments may contribute to emotional dysregulation in posters and viewers. Images may spread suicidal thoughts and lead to mimicking of acts *Protective impact on social connection*. 36.6% of responses showed care, empathy, gave advice, with some viewers calling for help (e.g. Police). The top 10 most influential microbloggers who reposted the original message showed concern and care for the person who self-harmed *Potential mechanism(s): Social positive reinforcement*. Study authors refer to hyperpersonal model where the consequence of caregiving responses may positively address and reinforce a suicidal individual's underlying goal of self-presentation (i.e. a cry for help)

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
8. Hetrick et al. (2020) Australia Qualitative Cross-sectional	Identify triggers to self-harm and helpful strategies to manage self-harm urges in young people	Graphic images on Tumblr or Instagram (e.g. cuts, scars)	Retrospective self-reported impact via semi-structured interviews	*Harmful impacts on self-harm, social comparison, and emotion.* Exposure to images was reported to normalise self-harm, was triggering, and led participants to self-harm. Participants compared their own self-harm, to the images of others' self-harm online, and led people feeling bad and more isolated if their self-harm was less severe as they felt they did not conform to social norms of self-harm. *Potential mechanism(s): Comparison and competition, descriptive norms, and emotional dysregulation.* Normalising influence of images and role of comparison, which triggered distressing emotions and self-harm
9. Jacob et al. (2017) Wales, UK Qualitative Cross-sectional	Explored young people's experiences of using the Internet for self-harm purposes and impact of self-harm images	Online self-harm images more generally (although participants recruited via Facebook and referred to Facebook & Tumblr)	Retrospective self-report via semi-structured interviews; 19 conducted individually, and one conducted in a dyad	*Harmful impacts on self-harm, engagement behaviours, social comparison, identity, social connection, emotion, cognition and physiology.* Images were used to trigger self-harm rituals; share and discover new self-harm practices; brought back memories of self-harm that enabled participants to envisage how they could self-harm; brought desire to recreate images or perform more severe self-harm; invoked a physical reaction; inspired behavioural enactment and competition to become 'better self-harmers'. Tumblr perceived to permit an immediate and intimate connection with others, which had led to normalisation and exacerbation of self-harm in participants. *Potential mechanism(s): Comparison and competition, descriptive norms, emotional dysregulation and regulation, priming, and social integration and regulation.* Normalising role of self-harm images through behavioural reinforcement or encouragement of self-harm enactment, sharing of techniques, and competition. Images evoke a physical reaction and inspire behavioural enactment
10. Jaroszewski et al. (2020)[a] USA Quantitative Cross-sectional Experimental (only Study 1 conducted online)	Validating first-person (FP) suicide attempt images and supplementary data analysis exploring whether those with suicidal history found suicidal images less	FP suicide images (e.g. looking down at a butcher's knife before it penetrates their abdomen) compared with positive (looking at a cookie), third person (TP) suicide	Questionnaire asked participants to rate images on valence (pleasantness), arousal, and threat	*Harmful impact on emotion.* Participants rated FP-suicide images as much more aversive, arousing and threatening than IAPS neutral images, and similar to both TP-suicide and IAPS negative images across all three rating dimensions. In the supplementary data analyses: For valence, the suicide group rated TP-Suicide images as significantly more pleasant and FP-positive images as less pleasant than the Nonsuicide

(continues)

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
	aversive than those with no suicidal history	image (someone jumping off a cliff), and other positive and negative images (using International Affective Picture System (IAPS))		group. For arousal, the suicide group rated FP-suicide, TP-suicide and IAPS negative images as significantly more arousing than the Nonsuicide group. In terms of threat, the Suicide group rated FP-suicide and TP-suicide images as significantly less threatening than the Nonsuicide group *Potential mechanism(s): Habituation and lower aversion.* Individuals with suicidal history may have lower aversion to these stimuli (Benefits & Barriers Model; Hooley & Franklin, 2018) and/or have acquired capability through habituation from repeated exposure (Joiner, 2005)
11. Lavis and Winter (2020) UK Qualitative Cross-sectional	Explored why and how young people engage with online self-harm content and the harms and benefits	Self-harm interactions including self-harm images and/or videos on Instagram, Reddit and Twitter	Explored impact through 10,169 original posts, 36,934 comments (written and pictorial content) and retrospective self-reported impact via 10 semi-structured interviews with young people who had used social media to engage with self-harm content	*Harmful impacts on self-harm, social connection and emotion.* Seeing images of others' self-harm could trigger viewers to self-harm due to their distress, not necessarily due to the image per se. To maintain online support, some viewers felt they needed to self-harm more severely to keep showing they needed support. Online peer-support may also preclude offline help-seeking, thinking others will not understand. Ideas for how-to self-harm safely were also offered in response to graphic images. *Potential mechanism(s): Assortative relating, descriptive norms, and emotional dysregulation.* Images may normalise or exacerbate self-harm; however, people who access this content likely to already be self-harming (i.e. assortative relating) *Protective impact on social connection.* The impact of people sharing self-harm imagery online was often peer concern and support. Peer support was a key driver for seeking online interactions related to self-harm. Potential mechanism(s): n/a
12. Lewis et al. (2012) Canada Quantitative Cross-sectional	Examined viewers' responses to NSSI YouTube videos to determine their	100 most-viewed YouTube Videos with NSSI content (50 videos with no live person, and 50 videos	Analysing 869 randomly selected comments from 100 most viewed NSSI videos on YouTube	*Harmful impacts on engagement behaviours, social connection, emotion and cognition.* Most frequent comments were self-disclosure of own NSSI experiences (38.39%), followed by feedback for video uploader, including admiration (15.40%) or

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
	potential risks and benefits	with at least one live person) whereby majority had graphic NSSI images		encouragement (11.15%). The majority did not mention recovery at all (42.89%) and indicated that they were still self-injuring (34.00%). Comments about sharing NSSI methods and strategies (e.g. wound cleaning, concealing) were not common (1.38%). Some comments mentioned that the video content was triggering and evoked NSSI urges (1.72%). *Potential mechanism(s): Assortative relating, descriptive norms, emotional dysregulation and social positive reinforcement*. Viewers' responses to videos by sharing own self-injury experiences or providing admiration/encouragement, may maintain NSSI behaviour through normalisation, and lack of recovery talk *Protective impacts on engagement behaviours and social connection*. Some individuals indicated they had recovered (7.11%) or expressed a desire to recover (3.78%). 2.41% of comments encouraged the uploader to seek help, and some offered to communicate with the uploader through YouTube or e-mail (1.61%). *Potential mechanism(s)*: n/a
13. Seko et al. (2015) Canada Qualitative Cross-Sectional	Explore what drives NSSI content creation and sharing, and how content creators react to others' NSSI content	Self-harm content online; 7/17 participants had posted images, and 3/17 had posted videos	Retrospective self-reported reactions explored via semi-structured email/text chat interviews	*Harmful impacts on self-harm, engagement behaviours, identity, social connection and emotion*. Some participants reported that viewing/reading NSSI materials could increase the urge to engage in NSSI, particularly if already in a negative state of mind. Some felt stronger attachment to textual representations as they provide greater context, while others found visual content more powerful and triggering. Some participants reported positive feelings when people commented or reblogged content. Online communities were seen to provide participants with emotional support that was largely absent from their offline life. Crowd mentality pushed others to create and upload similar images and advocate for NSSI community *Potential mechanism(s): Assortative relating, emotional dysregulation, social positive reinforcement, and social learning*. NSSI content could trigger NSSI in users with a history of self-harm when in a vulnerable emotional state *Protective impacts on self-harm, social comparison, identity, social connection, emotion and cognition.*

(continues)

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
				Thematic analysis of participants' narratives identified two motives: *self-oriented motivation* (to express self and creativity, to reflect on NSSI experience, to mitigate self-destructive urges); and *social motivation* (to support similar others, to seek out peers, to raise social awareness). *Potential mechanism(s): Emotional regulation, disgust, social positive reinforcement and self-reflection*. NSSI content could deter NSSI (through vicarious relief and repugnance). Online social connection, validation, emotional support and reflection can lead to NSSI urge regulation in vulnerable users with history of self-harm
14. Seko and Lewis (2018) Canada Quantitative Cross-sectional	Examined visual, textual and compositional attributes of photo-based images on Tumblr, exploring how they represent self-injury (SI) and how images are viewed	294 posts on Tumblr containing self-harm images (e.g. direct self-injury or memes from the media). 123 images (41.8% of the total sample) depicted self-injured bodies or body parts (mostly cuts)	Explored impact through 294 posts and 11,262 reblogs and attached narratives (texts, tags, notes) of self-harm images	*Harmful impact on engagement behaviours*. The most reblogged among direct self-injury images were denotative images (e.g. close-up shots of wounds; $n = 13$, 590 reblogs), followed by SI normalisation images, which normalised SI ($n = 12$, 137 reblogs). In 20 images (15 Direct SI images and 5 Indirect SI images) showing self-inflicted wounds (mostly open, bleeding wounds), razor blades, or other materials that symbolise SI, the act of SI was presented as an effective coping mechanism. *Potential mechanism(s): Descriptive norms and social integration and regulation*. Narrative reinforcement (normalisation/descriptive norms) of self-harm images and ideas, and context of reblogging, may maintain self-harm in vulnerable viewers who identify with these messages *Protective impacts on engagement behaviours and identity*. The third most reblogged among direct self-injury images included hopeful/pro-recovery images ($n = 21$, 132 reblogs), depicting scars/healing wounds which can signify recovery and resilience *Potential mechanism(s): Descriptive norms, social positive reinforcement, and social integration and regulation*. Images may promote recovery narratives
15. Sternudd (2012) Authors from Sweden (respondents mostly USA and UK)	Examined participants' reasons for producing and looking at images of self-injury, as well as	Online images of wounds and scars	Retrospective self-report reactions to self-harm images elicited via questionnaires	*Harmful impacts on self-harm, engagement behaviours, social comparison, identity, social connection, emotion and cognition*. 25% of statements were interpreted as negative impacts (could inspire/admire, trigger, give

Table 2 (continued)

Author (year) Country Design	Aims	Image source and type	How impact of self-harm images was measured	Potentially harmful or protective impacts/responses to viewing self-harm images and associated potential mechanisms explicitly or implicitly discussed by study authors
Mixed-Methods Cross-sectional (closed questions analysed quantitatively; open questions analysed qualitatively)	their reactions to their self-harm images and others' self-harm images			tips, lead to comparison and competition, and exacerbate self-harm). Impact of images varied depending on mood at the time, and images were used to promote the right mood for self-harm. Those with less experience of self-harm or more frequent current self-harm were more negative about effects of exposure to self-harm images *Potential mechanism(s): Assortative relating, comparison and competition, descriptive norms, and emotional dysregulation and regulation*. Images may normalise and trigger self-harm and impact on identity as a 'self-injurer', depending on individual and situational factors *Protective impacts on self-harm, social comparison, identity, social connection, emotion and cognition*. 40% of statements were interpreted as positive impact. Images could alleviate urges or deter participants from self-harm. Photos were published as a form of help-giving (e.g. deterrent), way of getting instant help, and evaluate the pros/cons. Individuals with longer self-harm history/former self-harmers were twice as likely to state they were positively influenced *Potential mechanism(s): Emotional regulation and self-reflection*. Images may alleviate self-harm through emotional regulation or self-reflection

Definitions of mechanisms: Assortative relating – individuals who possess similar difficulties/interests may be more likely to form relationships; Aversion – something that arouses strong feelings of dislike (potentially a natural barrier that may protect some people from self-harm); Descriptive norms (normalisation) – perception that a behaviour is more common than it might be (e.g. self-harm or recovery); Disgust – something that arouses revulsion (potentially a natural barrier that may protect some people from self-harm); Emotional dysregulation – when negative affect is amplified; Emotional regulation – when negative affect is reduced and/or positive affect is increased; Habituation – reduction in aversion through repeated exposure; Social positive reinforcement – individuals engage in behaviours to generate desired social outcomes (e.g. connection and validation); Priming – activation of one thought may trigger other thoughts/images or behaviours; Social comparison – comparing self with others; Social integration and regulation – sharing of ideas in closed/cohesive communities; Social learning – modelling and imitation of behaviour especially if there are perceived shared characteristics.
NR, not reported; NSSI, nonsuicidal self-harm; SI, self-injury.
[a]Data from study 1 and/or supplementary info only.

sectional, with the exception of Arendt, Scherr, and Romer (2019) who used a longitudinal design. The impact of exposure to self-harm images was studied through questionnaires (Arendt et al., 2019; Sternudd, 2012), semi-structured qualitative interviews (Brown, Fischer, Goldwich, & Plener, 2020; Chen et al., 2021; Hetrick et al., 2020; Jacob et al., 2017; Lavis & Winter, 2020; Seko, Kidd, Wiljer, & McKenzie, 2015), analysis of comments alongside images or patterns of posting and reposting/reblogging of images (Baker & Lewis, 2013; Brown et al., 2018; Fu, Cheng, Wong, & Yip, 2013; Lavis & Winter, 2020; Lewis et al., 2012; Seko & Lewis, 2018), or by measuring exposure to self-harm images directly in an experimental design (Cha et al., 2016; Jaroszewski, Kleiman, Simone, & Nock, 2020). The total number of participants recruited or comments/posts analysed varied depending on the study design (i.e. interviews: n = 188 participants; longitudinal: n = 729; experimental: n = 3,535; naturalistic observational: n = 4,938 and 11,332 posts/comments). Four studies originated from Canada, three from the USA, two each from China, Germany and the UK, and one each from Australia and Sweden. Social media platforms studied included Instagram, Reddit, Tumblr, Twitter, San Weibo (China), and YouTube, as well as online self-harm forums and the internet more broadly. In Jacob et al's (2017) study, recruitment was restricted to participants with a Facebook account, but they did not specifically inquire about exposure to imagery on Facebook per se, asking instead about the Internet more generally. It is, therefore, not known to what extent participants responded with reference to Facebook. Overall, there was a lack of generalisability (e.g. the majority of participants were White and female). Age was reported in nine studies, with most participants aged between 14 and 27 years old (eight studies), but the participants in one study had a mean age of 39 years (Jaroszewski et al., 2020). Demographic characteristics of participants were not reported in two studies (Baker & Lewis, 2013; Lewis et al., 2012) and there was minimal demographic information in three studies due to the anonymous nature of posts (Fu et al., 2013; Lavis & Winter, 2020; Seko & Lewis, 2018). Demographic information was only reported for a subset of participants in two studies (Brown et al., 2018; Lavis & Winter, 2020). Eleven studies included information indicating that participants had a history of self-harm or suicidality, or were currently self-harming (Arendt et al., 2019; Baker & Lewis, 2013; Brown et al., 2020; Chen et al., 2021; Hetrick et al., 2020; Jacob et al., 2017; Jaroszewski et al., 2020; Lavis & Winter, 2020; Lewis et al., 2012; Seko et al., 2015; Sternudd, 2012). Little was known about participants' mental health history as this was not, or could not, be examined. Participants in Jacob et al's (2017) study most frequently

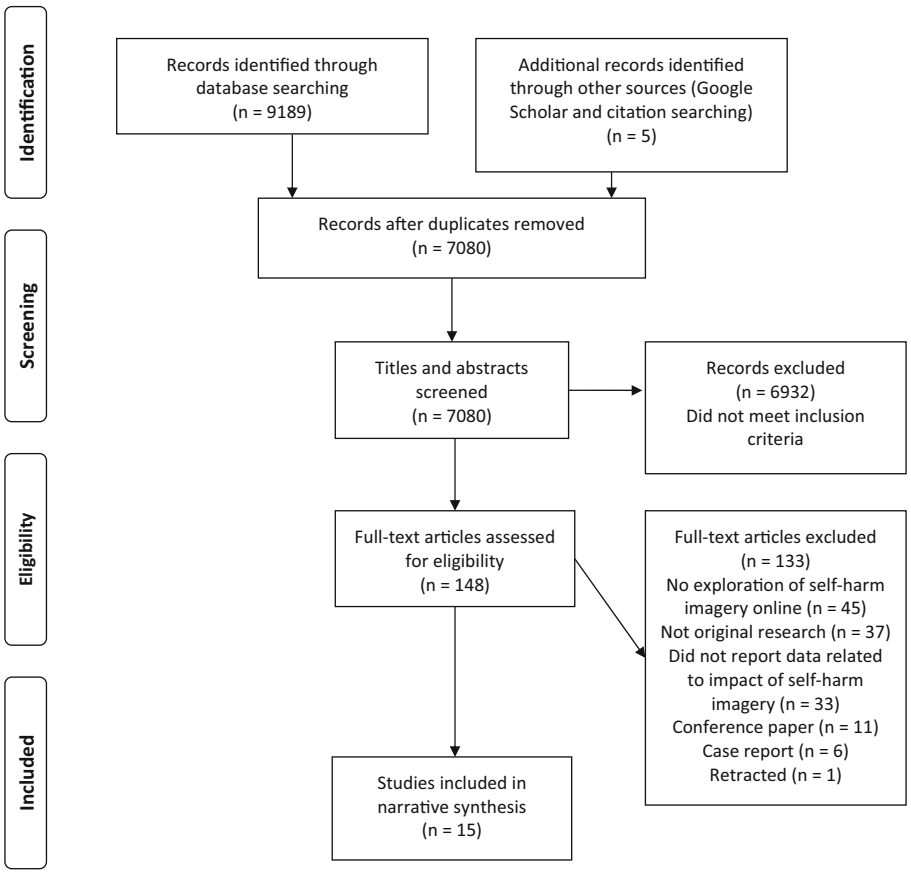

Figure 1 PRISMA diagram

made reference to Tumblr, but one participant referred to Facebook. Most studies explored only one online platform, but Lavis and Winter (2020) investigated Instagram, Reddit, and Twitter, and Hetrick et al. (2020) reported results relating to the impact of graphic images on Instagram and Tumblr.

Eight studies were rated as high quality, five as medium quality and two as low quality (Table S3). Causality and directionality of impact was not determined in any study. Most studies were cross-sectional and relied on retrospective self-report of exposure to self-harm images and impact. Also, most studies did not directly measure exposure to self-harm images (type or severity) or demographic and vulnerability factors (e.g. exposure to self-harm via other sources, such as the media or family/friends), with the exception of Arendt et al. (2019) who assessed previous exposure to self-harm through images on Instagram, media, and via friends and family. Some studies could not measure exposure due to methodological limitations associated with naturalistic observations of comments/posts. The longer-term duration of impact of exposure to self-harm images was not measured in any study. When measures were used these were often nonvalidated single-item measures. Interpretation of findings from observational studies (e.g. comments) were subjective and open to bias, and not all authors acknowledged or controlled for this. Additionally, studies analysing posted material focused on the assessment of the potential impact on active users who left comments. The perspectives of individuals who viewed but did not interact with self-harm images were not usually assessed.

Fourteen studies discussed the potential impact of 'graphic' self-injurious images (images and/or videos depicting cuts) and one explored self-injurious- and suicidal-related images (e.g. first-person images, such as looking down at a knife or gun barrel, and third-person images, such as images depicting a person potentially attempting suicide from an observer perspective; Jaroszewski et al., 2020).

Only two quantitative studies examined evidence for a contagion mechanism (Arendt et al., 2019; Brown et al., 2018), although other study authors did sometimes discuss possible mechanisms explicitly or implicitly (see Table 3; Table S1). Arendt et al. (2019) found that intentional and accidental exposure to self-harm images on Instagram predicted self-harm and suicide-related outcomes one month later. In contrast, Brown et al.s' (2018) time-related analysis did not find any evidence of social contagion (although they did find evidence of social reinforcement) as they did not identify any clustering of nonsuicidal self-harm (NSSI) photos on Instagram. However, their analysis was limited to posted self-harm images, and contagion effects may occur over a longer timeframe, and also offline.

A summary of the potentially harmful and protective impacts and mechanisms associated with viewing online images of self-harm is shown in Table 3 (Tables S1, S5 and S6).

Potentially harmful impacts of viewing online self-harm images

All studies found harmful impacts of viewing self-harm images online. Most evidence for the potentially harmful impacts related to: self-harm behaviours, engagement behaviours (e.g. sharing or commenting on images), social comparison (Festinger, 1954) which encourages competitiveness relating to self-harm, and promotion of a self-harmer identity and social connection that perpetuates or escalates self-harm, as well as emotional dysregulation, priming of self-harm cognitions (Berkowitz, 1984), and physiological responses increasing self-harm urges and self-harm.

Self-harm behaviour. Eight studies reported that exposure to graphic images or videos could normalise and 'trigger' self-harm and increase its frequency and/or severity (Arendt et al., 2019; Baker & Lewis, 2013; Brown et al., 2020; Hetrick et al., 2020; Jacob et al., 2017; Lavis & Winter, 2020; Seko et al., 2015; Sternudd, 2012). Arendt et al. (2019) found that a third (32.5%) of participants in their study carried out the same or similar type of self-harm after seeing self-harm on Instagram, suggesting a social learning (imitative) influence. Some participants reported intentionally searching for self-harm images (Arendt et al., 2019), sometimes to trigger the right mood to self-harm (Jacob et al., 2017; Sternudd, 2012), or if they were already in a negative state of mind, self-harm images could potentiate self-harm (Seko et al., 2015). Some authors (e.g. Arendt et al., 2019; Lavis & Winter, 2020) suggested a possible assortative relating influence, in that individuals who possess similar difficulties/interests may be more likely to form relationships and seek out others (i.e. those who self-harm) online, as most participants or users (i.e. those viewing, posting and/or engaging with online images) were usually already self-harming. Those who shared their self-harm images online appeared to engage in more severe self-harm, sometimes encouraged by the online self-harm community (Jacob et al., 2017).

Engagement with self-harm images, social comparison and identity. Nine studies reported that viewing self-harm images could encourage self-harm as a result of others sharing, and commenting on the images. Viewing images encouraged some individuals to start self-harming or to post their self-harm images, portraying self-harm as desirable and potentially normalising these behaviours (e.g. 'crowd mentality') (Baker & Lewis, 2013; Brown et al., 2020; Jacob et al., 2017; Seko et al., 2015). Images depicting self-harm wounds or more severe self-harm elicited more comments or attention than images

Table 3 Summary of potentially harmful and protective impacts and associated mechanisms

		Areas of potentially harmful and protective impacts with associated mechanisms ([a]explicitly or implicitly discussed by study authors; [b]identified by review authors)							
		Self-harm behaviours	Engagement behaviours	Social comparison	Identity	Social connection	Emotion	Cognition	Physiology
1. Arendt et al. (2019)	Harmful	[a]Assortative relating [a]Social learning [b]Descriptive norms	n/a	n/a	n/a	n/a	[a]Emotional dysregulation	[b]Priming (including mental imagery)	n/a
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
2. Baker and Lewis (2013)	Harmful	[a]Descriptive norms [b]Assortative relating [b]Social positive reinforcement [b]Social integration and regulation [b]Social learning	[a]Descriptive norms [b]Assortative relating [b]Social positive reinforcement [b]Social integration and regulation [b]Social learning	[a]Comparison and competitiveness	n/a	[a]Descriptive norms [b]Assortative relating [b]Social positive reinforcement [b]Social integration and regulation [b]Social learning	n/a	[b]Priming (including mental imagery)	n/a
	Protective	[a]Aversion (barrier) [a]Emotion regulation or automatic reinforcement	n/a	n/a	n/a	[b]Assortative relating [b]Descriptive norms [b]Social positive reinforcement [b]Social integration and regulation	[a]Aversion (barrier) [a]Emotional regulation or automatic reinforcement [b]Assortative relating [b]Descriptive norms [b]Social positive reinforcement [b]Social integration and regulation	[a]Emotion regulation or automatic reinforcement	[a]Emotion regulation or automatic reinforcement
3. Brown et al. (2018)	Harmful	n/a	[a]Social positive reinforcement [b]Descriptive norms [b]Social integration and regulation	n/a	n/a	[a]Social positive reinforcement [b]Descriptive norms [b]Social integration and regulation	n/a	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	[a]Social positive reinforcement	n/a	n/a	n/a
4. Brown et al. (2020)	Harmful	[a]Emotional dysregulation [a]Social learning	[a]Social positive reinforcement [a]Social learning [b]Assortative relating [b]Descriptive norms	[b]Comparison and competitiveness	n/a	[a]Social positive reinforcement [a]Social learning [b]Assortative relating [b]Descriptive norms	[a]Emotional dysregulation [a]Social learning	n/a	n/a
	Protective	[a]Disgust (barrier)	n/a	n/a	[b]Self-reflection	[a]Social positive reinforcement [b]Social integration and regulation	[a]Disgust (barrier)	n/a	n/a
5. Cha et al. (2016)	Harmful	n/a	n/a	n/a	n/a	n/a	[a]Emotional dysregulation	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
6. Chen et al. (2021)	Harmful	n/a	n/a	n/a	n/a	n/a	n/a	[a]Social learning [b]Priming	n/a
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 3 (continued)

		Areas of potentially harmful and protective impacts with associated mechanisms ([a]explicitly or implicitly discussed by study authors; [b]identified by review authors)							
		Self-harm behaviours	Engagement behaviours	Social comparison	Identity	Social connection	Emotion	Cognition	Physiology
7. Fu et al. (2013)	Harmful	n/a	[a]Social positive reinforcement [b]Assortative relating [b]Descriptive norms [b]Social integration and regulation	n/a	n/a	n/a	[a]Emotional dysregulation	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	[a]Social positive reinforcement	n/a	n/a	n/a
8. Hetrick et al. (2020)	Harmful	[a]Descriptive norms [a]Emotional dysregulation [b]Assortative relating [b]Social learning	n/a	[a]Comparison and competitiveness	n/a	n/a	[a]Emotional dysregulation	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
9. Jacob et al. (2017)	Harmful	[a]Descriptive norms [a]Emotional dysregulation [b]Assortative relating [b]Social learning	[a]Descriptive norms [b]Social positive reinforcement	[a]Comparison and competitiveness	[a]Social integration and regulation [b]Assortative relating [b]Social positive reinforcement	[a]Descriptive norms [a]Social integration and regulation [b]Assortative relating [b]Social positive reinforcement [b]Social learning	[a]Emotional dysregulation [a]Emotional regulation or automatic reinforcement	[a]Priming (including mental imagery) [b]Social learning	[a]Emotional dysregulation [b]Priming
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
10. Jaroszewski et al. (2020)	Harmful	n/a	n/a	n/a	n/a	n/a	[a]Habituation [a]Less aversion (lower barrier)	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
11. Lavis and Winter (2020)	Harmful	[a]Assortative relating [a]Descriptive norms [b]Social integration and regulation [b]Social learning	n/a	n/a	n/a	[a]Assortative relating [a]Descriptive norms [b]Social positive reinforcement [b]Social integration and regulation [b]Social learning	[a]Emotional dysregulation	n/a	n/a
	Protective	n/a	n/a	n/a	n/a	[b]Positive social reinforcement	n/a	n/a	n/a
12. Lewis et al. (2012)	Harmful	n/a	[a]Assortative relating [a]Descriptive norms [a]Social positive reinforcement	n/a	n/a	[a]Assortative relating [a]Descriptive norms [a]Social positive reinforcement [b]Social integration and regulation	[a]Emotional dysregulation	[b]Priming	n/a
	Protective	n/a	[b]Descriptive norms [b]Social positive reinforcement	n/a	n/a	[b]Descriptive norms [b]Social positive reinforcement	n/a	n/a	n/a
13. Seko et al. (2015)	Harmful	[a]Assortative relating [a]Emotional dysregulation [a]Positive social reinforcement	[a]Assortative relating [a]Social positive reinforcement [a]Social learning	n/a	[a]Assortative relating [a]Social positive reinforcement [a]Social learning	[a]Assortative relating [a]Social positive reinforcement [a]Social learning [b]Descriptive norms	[a]Emotional dysregulation	n/a	n/a

(continues)

Table 3 (continued)

		Areas of potentially harmful and protective impacts with associated mechanisms ([a]explicitly or implicitly discussed by study authors; [b]identified by review authors)							
		Self-harm behaviours	Engagement behaviours	Social comparison	Identity	Social connection	Emotion	Cognition	Physiology
		[a]Social learning [b]Descriptive norms	[b]Descriptive norms [b]Social integration and regulation		[b]Descriptive norms [b]Social integration and regulation	[b]Social integration and regulation			
	Protective	[a]Emotional regulation or automatic reinforcement [b]Disgust (barrier)	n/a	[a]Disgust (barrier) [b]Self-reflection [b]Comparison	[b]Self-reflection	[a]Social positive reinforcement [b]Assortative relating [b]Descriptive norms [b]Social integration and regulation	[a]Disgust (barrier) [a]Emotional regulation or automatic reinforcement	[a]Emotional regulation or automatic reinforcement [a]Self-reflection [b]Priming	n/a
14. Seko and Lewis (2018)	Harmful	n/a	[a]Descriptive norms [a]Social integration and regulation [b]Social positive reinforcement	n/a	n/a	n/a	n/a	n/a	n/a
	Protective	n/a	[a]Descriptive norms [a]Social positive reinforcement [a]Social integration and regulation	n/a	[b]Self-reflection	n/a	n/a	n/a	n/a
15. Sternudd (2012)	Harmful	[a]Assortative relating [a]Descriptive norms [a]Emotional dysregulation [b]Habituation [b]Re-experiencing emotional mental imagery [b]Social integration and regulation	[a]Descriptive norms	[a]Comparison and competitiveness	[a]Emotional dysregulation	[a]Descriptive norms [b]Social positive reinforcement [b]Social integration and regulation	[a]Assortative relating [a]Emotional dysregulation [a]Emotional regulation [b]Habituation [b]Re-experiencing emotional mental imagery	[b]Priming (including mental imagery)	n/a
	Protective	[a]Emotional regulation or automatic reinforcement [b]Aversion (barrier) [b]Self-reflection	n/a	[b]Aversion (barrier) [b]Emotional regulation or automatic reinforcement [b]Self-reflection [b]Social integration and regulation	[b]Self-reflection	[b]Descriptive norms [b]Social positive reinforcement	[a]Emotional regulation or automatic reinforcement [b]Aversion (barrier) [b]Re-experiencing emotional mental imagery	[a]Emotional regulation or automatic reinforcement [a]Self-reflection [b]Aversion (barrier)	n/a

Definitions of mechanisms: Assortative relating – individuals who possess similar difficulties/interests may be more likely to form relationships; Aversion – something that arouses strong feelings of dislike (potentially a natural barrier that may protect some people from self-harm); Descriptive norms (normalisation) – perception that a behaviour is more common than it might be (e.g. self-harm or recovery); Disgust – something that arouses revulsion (potentially a natural barrier that may protect some people from self-harm); Emotional dysregulation – when negative affect is amplified; Emotional regulation – when negative affect is reduced and/or positive affect is increased; Habituation – reduction in aversion through repeated exposure; Social positive reinforcement – individuals engage in behaviours to generate desired social outcomes (e.g. connection and validation); Priming – activation of one thought may trigger other thoughts/images or behaviours; Social comparison – comparing self with others; Social integration and regulation – sharing of ideas in closed/cohesive communities; Social learning – modelling and imitation of behaviour especially if there are perceived shared characteristics.

without self-harm wounds/scars or those with less severe wounds (Brown et al., 2018, 2020). Publishing and reblogging images, and commenting on images, often normalised self-harm as an effective coping strategy (Lewis et al., 2012; Seko & Lewis, 2018; Sternudd, 2012). Users also positively reinforced the sharing of images by showing admiration, complimenting, or encouraging uploaders to create and/or upload similar images (Brown et al., 2018; Lavis & Winter, 2020; Lewis et al., 2012; Seko et al., 2015), and by showing empathy and offering help (Brown et al., 2018; Fu et al., 2013).

Five studies reported that individuals would compare images to their own self-harm and reported a sense of pride if their self-harm appeared worse (Brown et al., 2020), or a sense of failure if their self-harm was not as severe (Hetrick et al., 2020; Jacob et al., 2017). Viewing images sometimes encouraged individuals to compete with others to become 'better self-harmers' and engage in more severe forms of self-harm (Baker & Lewis, 2013; Jacob et al., 2017; Sternudd, 2012), as well as to create and share self-harm images, which could encourage the development of a self-harm identity and lead to an escalation in self-harm (Jacob et al., 2017; Sternudd, 2012). Some individuals felt a desire to advocate for self-harm communities, sharing 'what's behind the scars' (Seko et al., 2015). Others also reported being triggered by viewing more severe self-harm or scars to their own, where their fading scars could be experienced as a loss of identity (Sternudd, 2012).

Social connection. Eight studies reported evidence that online images facilitated immediate connection, validation, help and support that may lead to unintended consequences (Baker & Lewis, 2013; Brown et al., 2018, 2020; Jacob et al., 2017; Lavis & Winter, 2020; Lewis et al., 2012; Seko et al., 2015; Sternudd, 2012). Individuals shared images/comments to help others feel less alone and portrayed self-harm as an acceptable coping mechanism (Baker & Lewis, 2013; Brown et al., 2020; Lewis et al., 2012; Seko et al., 2015; Sternudd, 2012). Individuals also shared ideas for self-harm methods or how to hide injuries (e.g. Brown et al., 2018; Jacob et al., 2017; Lavis & Winter, 2020; Lewis et al., 2012; Sternudd, 2012), and where on the body to self-harm 'safely' (Lavis & Winter, 2020), potentially normalising, validating and maintaining self-harm (Baker & Lewis, 2013; Jacob et al., 2017; Lavis & Winter, 2020; Sternudd, 2012). To maintain social connection and support, one participant described how their self-harm worsened as they needed to keep showing others their need for help, which might perpetuate, or even escalate, self-harm in order to maintain the same level of peer-support (Lavis & Winter, 2020). Social support may also inadvertently preclude offline help-seeking (Baker & Lewis, 2013), with Lavis and Winter (2020) finding that individuals thought 'normies' (those who do not self-harm) would not understand. There was little or no mention of successful referrals for professional help, or of recovery (Brown et al., 2020; Lewis et al., 2012), with 34% of comments to videos in one study indicating that individuals were still self-harming (Lewis et al., 2012).

Emotional impacts. A change in emotions increasing the likelihood of self-harm associated with viewing images was reported in 11 studies. Graphic images and videos were found to be emotionally disturbing, and potentially triggering of self-harm behaviour (Arendt et al., 2019; Brown et al., 2020; Jacob et al., 2017; Jaroszewski et al., 2020; Lavis & Winter, 2020; Lewis et al., 2012; Seko et al., 2015; Sternudd, 2012). Images could lead to changes in affect, such as depression, grief, shock, disgust, isolation, hostility, empathy, or concern (Cha et al., 2016; Fu et al., 2013; Hetrick et al., 2020; Lavis & Winter, 2020; Lewis et al., 2012; Sternudd, 2012), and be intentionally used to induce the right mood for self-harm (Jacob et al., 2017; Sternudd, 2012). Cha et al. (2016) found that viewing self-harm images led to a greater decline in mood compared to words, particularly in females, suggesting images may have a greater impact on emotions. Seko et al. (2015) also described how several participants in their study reported that images were more powerful than textual representations (although two participants reported the opposite). Jacob et al. (2017) found that images rather than textual interactions were the primary reason cited for using the Internet for self-harm purposes (almost 75% of participants), as images were described as more impactful than words. Tumblr, in particular, permitted the anonymous sharing of images, and was the preferred platform.

Study participants with less experience of self-harm or suicidal behaviours were more likely to describe the impact of viewing self-harm images as negative (Jaroszewski et al., 2020; Sternudd, 2012). Jaroszewski et al. (2020) found that individuals who had experienced suicidal thoughts or behaviours in their lifetime rated third-person suicide images as more pleasant, and third-person and first-person suicide images as more arousing and less threatening than those who had not experienced these thoughts or behaviours themselves. The authors suggested that this might result from habituation, where repeated exposure to this material becomes less aversive over time, or because psychological barriers, which might typically inhibit self-harm or suicide, are lower (Hooley & Franklin, 2018).

Cognitive impacts. Six studies reported that self-harm images increased self-harm cognitions, such as urges or inspiration to self-harm (Baker & Lewis, 2013; Chen et al., 2021; Jacob et al., 2017;

Lewis et al., 2012; Sternudd, 2012), and were positively associated with increased suicidal ideation and hopelessness, and negatively associated with reasons for living (Arendt et al., 2019). There was some evidence that viewing images evoked mental imagery (e.g. past self-harm), enabling individuals to imagine how it would feel to self-harm in the same way (Arendt et al., 2019; Sternudd, 2012), and increasing the intention to self-harm in a similar way (Jacob et al., 2017).

Physiological impacts. One study found evidence that images could evoke powerful physiological reactions (e.g. 'rush' and 'fast heart rate') which could trigger self-harm (Jacob et al., 2017).

Potentially protective impacts of viewing online self-harm images

Nine studies reported potentially protective impacts, with most of the evidence relating to self-harm behaviour, engagement behaviours, social comparison, identity, social connection and support, as well as emotional regulation and dysregulation, cognitive and physiological impacts mitigating self-harm urges and behaviours.

Self-harm behaviour. Four studies provided evidence that viewing self-harm images could prevent further self-harm, at least in the short-term (Baker & Lewis, 2013; Brown et al., 2020; Seko et al., 2015; Sternudd, 2012). Baker and Lewis (2013) found that viewing images gave participants a similar feeling to when they themselves self-harmed, such that the act of viewing images was a proxy for self-harm. This may suggest an emotional regulation mechanism. In other studies, viewing severe self-harm or seeing others' scars discouraged individuals from self-harming (Baker & Lewis, 2013; Brown et al., 2020; Seko et al., 2015; Sternudd, 2012), potentially through feelings of aversion or disgust as reported by participants (although this was not formally measured by study authors). Some participants used images strategically to avoid more severe self-harm acts (Baker & Lewis, 2013; Seko et al., 2015; Sternudd, 2012).

Engagement with self-harm images, social comparison and identity. Seko and Lewis (2018) found that the third most reblogged self-injury images included messages of hope and recovery, with 32 of 86 images depicting scars or healing wounds, rather than bleeding injuries, promoting a recovery narrative. Comments on images sometimes encouraged uploaders to seek help, or shared that they had stopped self-injuring themselves, or wanted to stop (Lewis et al., 2012).

Images depicting more severe self-harm than an individual's own acts sometimes served as a deterrent to them engaging in self-harm, perhaps due to self-reflection or aversion (Seko et al., 2015; Sternudd, 2012).

Four studies also reported a possible protective impact of viewing, creating or sharing self-harm images in terms of identity, allowing self-expression or promoting recovery narratives (Brown et al., 2020; Seko et al., 2015; Seko & Lewis, 2018; Sternudd, 2012). Self-reflection (as discussed in Ryan-Vig, Gavin, & Rodham, 2019) may enable individuals to change their relationship to self-harm (e.g. scars being seen as proof of healing, recovery and resilience), potentially leading to reduced self-harm and feeling more in control (Seko et al., 2015; Seko & Lewis, 2018; Sternudd, 2012).

Social connection. Eight studies reported findings suggesting that online activity could be protective for individuals. The posting of images encouraged online peer-support that could be missing offline, helping individuals to identify with others through self-disclosure, feeling more socially connected and validated and/or less isolated (Baker & Lewis, 2013; Brown et al., 2018, 2020; Lavis & Winter, 2020; Lewis et al., 2012; Seko et al., 2015; Sternudd, 2012), and mitigate urges to self-harm (Seko et al., 2015). Users sometimes made comments which encouraged others to stop self-harming (Brown et al., 2018), and/or posted images/comments to get help themselves or encourage others to seek help (Lewis et al., 2012; Sternudd, 2012), or offered help (Brown et al., 2018, 2020; Fu et al., 2013; Lewis et al., 2012; Sternudd, 2012). For instance, images of scars or self-harm were sometimes shared to discourage others from self-harming (Baker & Lewis, 2013; Sternudd, 2012).

Emotional impacts. Four studies reported that viewing images could lead to a change in emotional state that reduced the likelihood of self-harm, such as by providing vicarious 'relief', 'comfort' or 'sense of calm', so self-harm could be avoided (Baker & Lewis, 2013; Seko et al., 2015; Sternudd, 2012). Fear or disgust also seemed to deter some participants from self-harming or carrying out more severe forms of self-harm (Baker & Lewis, 2013; Brown et al., 2020; Seko et al., 2015).

Cognitive impacts. Three studies reported that anticipatory/vicarious relief in response to viewing self-harm images could alleviate urges to self-harm as if individuals had themselves self-injured (Baker & Lewis, 2013; Seko et al., 2015; Sternudd, 2012).

Physiological impact. There was some evidence that viewing images could cause a similar 'adrenaline rush' to cutting, thereby reducing self-harm behaviour (Baker & Lewis, 2013).

Discussion

Main findings

We identified 15 studies which reported evidence of the potential impacts of viewing self-harm images online. The methodological limitations of the study designs precluded the possibility of drawing causal conclusions, highlighting the infancy of research in this field. All studies found potentially harmful effects of exposure to images of self-harm, such as individuals being 'triggered' emotionally and cognitively, which could lead to self-harm, being encouraged to share images or compete with others, escalating and normalising self-harm as an acceptable coping response, through sharing tips and ideas, and the potential development of a 'self-harmer' identity, with little mention of recovery or offline help-seeking. Most studies found harmful impacts on emotions and engagement behaviours through sharing and commenting on self-harm images.

Nine studies reported potentially protective effects, such as the viewing of online images serving as a proxy for self-harm, being deterred from self-harm after viewing images, enabling self-reflection, being offered help and support, and feeling connected to others. The most consistent evidence for a protective impact related to social connection.

Viewing self-harm images could be both harmful and protective, depending on individuals' mood at the time (Seko et al., 2015; Sternudd, 2012). The heterogeneity of the findings suggests that images may have different impacts or functions (such as emotional regulation, dysregulation, social connection and belonging) at different times depending on the individual, and contextual and dynamic aspects of impacts and mechanisms.

The potential mechanisms associated with harmful and protective effects of viewing self-harm images were not investigated or explicitly referred to in most studies beyond ideas of contagion, which assumes causality. While both Arendt et al. (2019) and Brown et al. (2018) explored a contagion hypothesis, where exposure to self-harm images may increase self-harm, their findings were somewhat contradictory.

The potential mechanisms for harmful effects most commonly discussed or implied by study authors included normalisation, assortative relating, social learning, social positive reinforcement and emotional dysregulation (Table 3). We also identified social integration and regulation (sharing ideas; Mueller and Abrutyn (2016)), cognitive priming (including mental imagery), habituation, lower aversion to self-harm stimuli and comparison and competitiveness as harmful potential mechanisms. The protective mechanisms towards recovery implied by the study authors were normalisation, social integration and regulation, social positive reinforcement, emotional regulation, aversion and self-reflection. As protective mechanisms, we also identified assortative relating, priming, and social comparison. Three studies (i.e. Cha et al., 2016; Jacob et al., 2017; Seko et al., 2015) found some evidence that images have a greater emotional impact than words, supporting previous literature emphasising the importance of imagery, including mental imagery (Holmes & Mathews, 2010; Winkielman & Gogolushko, 2018).

The mechanisms identified in this review may provide insight into the motivations involved in viewing self-harm images online, such as emotional regulation, emotional dysregulation and social connection. This supports broader theories relating to the function of self-harm, for example, the four-function model (Bentley, Nock, & Barlow, 2014; Nock & Prinstein, 2004), which proposes that non-suicidal self-injury (NSSI) is maintained through four reinforcement mechanisms related to positive or negative automatic reinforcement (emotional regulation), and positive and negative social reinforcement. The findings also support other emotional regulation or interpersonal influence research and models (Hasking, Whitlock, Voon, & Rose, 2017; Hooley & Franklin, 2018; Klonsky, 2007; Rodham, Hawton, & Evans, 2004), social learning theories (Nock, 2009) and Joiner's (2005) interpersonal theory of suicide, whereby 'thwarted belonginess' and 'perceived burdensomeness' may drive individuals to seek support online. There was stronger evidence for potential impacts than there was for potential mechanisms. Unlike the suicide clustering and influence of media literature, there has been very little discussion of mechanisms in relation to the viewing of self-harm images. Further research investigating both impacts and mechanisms is required.

The participants in 11 of the studies had a reported history of self-harm and/or suicidality or were currently self-harming, so may have chosen to access online self-harm images due to pre-existing vulnerabilities, as suggested in relation to an assortative relating mechanism (Joiner, 2003), or through habituation to aversive stimuli over time (Joiner, 2005; see Table S1). This highlights the importance of intention and pre-existing vulnerability factors in determining the potential impacts of self-harm images.

Strengths and limitations of included studies

Surprisingly, we were only able to identify 15 studies that included empirical data related to the impacts of viewing self-harm images online. Most of the qualitative studies were of high quality, and most quantitative studies included large samples, but were of low or medium quality. Studies often reported inconsistent findings (see Tables 2 and 3), which might be explained by the heterogeneity of methodologies and because many studies only examined

one internet platform. Causality or directionality of impact could not be determined as most studies used cross-sectional and retrospective self-report measurement of exposure to self-harm images and assessed impact in participants who were already self-harming, with limited use of validated measures. Further limitations to the generalisability of findings were that most participants were White, Western and female.

Strengths and limitations of this review

Our findings are broadly consistent with those of similar reviews. Particular strengths of the review relate to being the first (to our knowledge) to explore both the potential impacts and the underlying mechanisms of viewing self-harm images online. It includes both quantitative and qualitative studies, and studies which analysed verbal content alongside images, as this is likely to influence how images are appraised. Care was also taken to separate mechanisms suggested (either explicitly or implicitly) by study authors from those identified by the review authors (see Table 3). Also, our inter-rater reliability for screening and quality analysis was high.

Grey literature was excluded, and only English-language publications were included, so some publication bias is possible. Studies were not excluded based on age of participants as this information was not always available due to the anonymous nature of some platforms. Whilst more young people are likely to be using these online spaces, impact may vary with age or over time through repeated exposure (Jaroszewski et al., 2020), so it was deemed important not to limit studies on this basis. It was unclear in some studies whether the identified impacts were protective, harmful, or both. Also, potential mechanisms involved were often not explored, so these were interpreted with consensus being reached between the review authors. Thus, the analysis of the themes relating to potential impact and mechanisms is subjective and not exhaustive. The findings are further limited in that not much is known about the impacts on those without a self-harm history, or those who unintentionally get exposed to such images. Given that most studies included participants with pre-existing vulnerabilities to self-harm/suicidality, it is possible that this may have influenced the results, whereby more harmful effects predominated.

Research implications

The study findings were used to inform a conceptual model of impacts of viewing self-harm images online for further testing (Figure 2). In this we suggest how harmful and protective impacts may arise through pre-existing vulnerability (e.g. self-harm history, previous exposure to images), contextual factors (e.g. mood, type of images, intentionally viewed or not), and appraisal of images by the viewer and others (e.g. through comments). These factors interact and change over time, suggesting that the impact of viewing self-harm images online may also change over time, such as following repeated exposure to images. Future research should investigate what contributes to the mixed findings related to harmful or protective effects, and in which contexts these occur, such as whether different platforms give rise to different impacts, possibly due to moderation or recovery narratives. Specifically, the potential impacts and mechanisms, including mediators and moderators, may vary between those with different self-harm or suicide histories. Future work should explore this by controlling for, or comparing, group differences related to self-harm exposure (e.g. non-suicidal), suicidality, mental health history and other variables of interest, as indicated in the potential vulnerability factors in Figure 2. Importantly, the impact of self-harm images on individuals who do not self-harm is unknown.

Given that the majority of the findings relate to White female participants, investigators conducting further research in this area should aim to recruit a more representative sample of participants with a range of genders, sexualities and ethnicities. Farooq et al. (2021) found that there was an increased proportion of presentations to hospital over time for minority ethnic children and adolescents compared with White ethnic groups, which emphasises the importance of considering cultural differences. Further research and interventions addressing the barriers to recruiting such participants to self-harm research is also warranted, as well as advertising research and engaging with individuals from different settings (e.g. primary care, schools, community and cultural groups).

Given that no study could determine directionality, causality, the duration of impact, or whether this changed over time, high-quality longitudinal methods (e.g. diary or experience sampling methodology) capable of capturing more timely and accurate information with less susceptibility to memory bias, should be used in future studies. Algorithm studies could also be conducted to evaluate the impact of promoting recovery-based narratives.

Clinical implications

Young people who self-harm are likely to use the internet in ways that increases their risk (Mitchell & Ybarra, 2007). Psycho-education for young people, parents/carers and professionals is important in order to enhance awareness and understanding of this potential for harm. Young people who present to services following self-harm should routinely be asked about access to online self-harm images. The eight areas of potential impact we have identified from the included studies could guide routine clinical assessments, by exploring the idiosyncratic

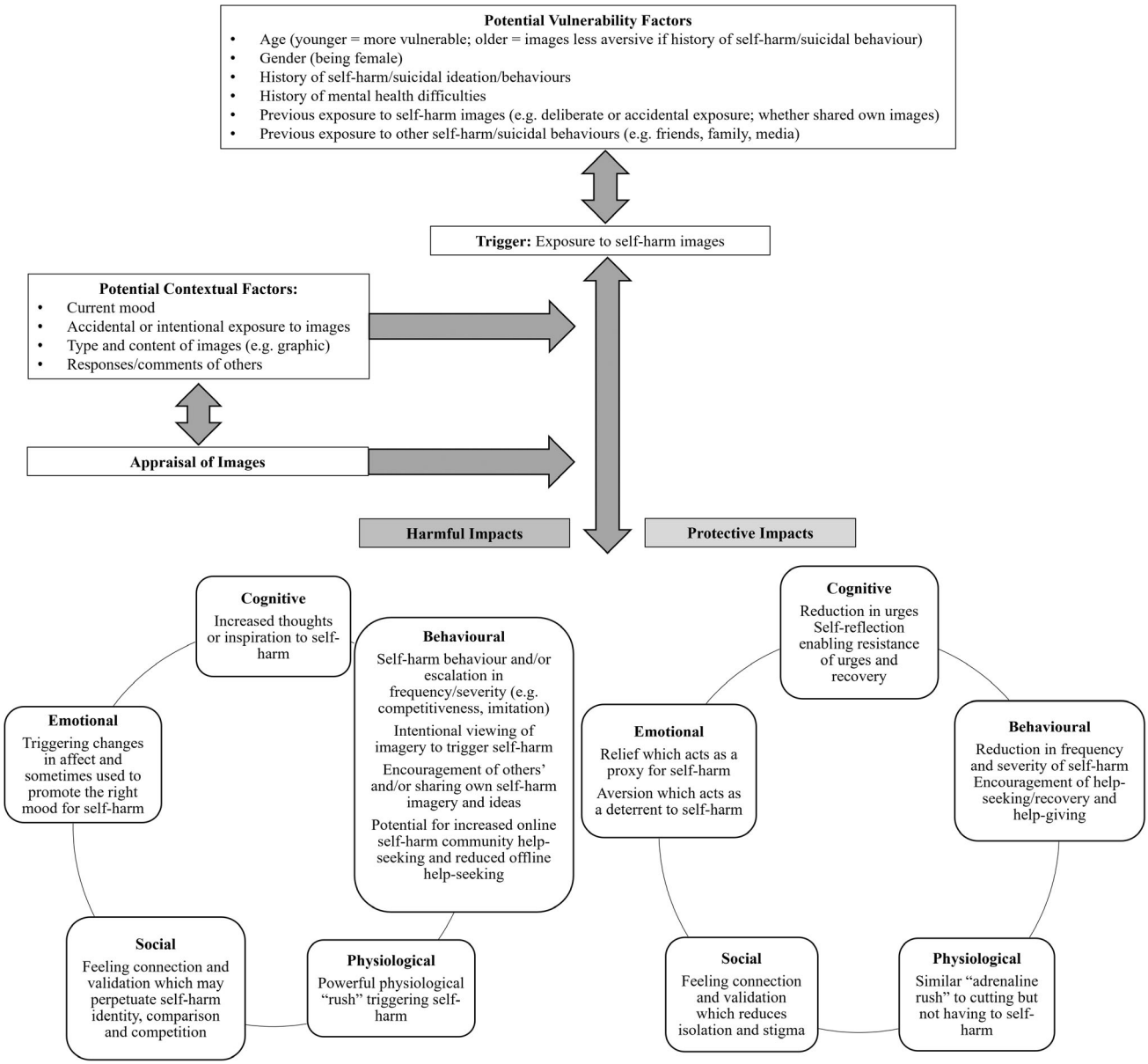

Figure 2 Model of impact related to viewing self-harm images online

impact, context and function of accessing self-harm images and how this may change over time (e.g. protective to harmful or vice versa) in order to identify relevant interventions and improve resilience, such as through developing emotional regulation skills and expanding offline peer and social support.

Most young people who self-harm do not access professional support (Kidger, Heron, Lewis, Evans, & Gunnell, 2012; Rowe et al., 2014), particularly among diverse ethnocultural groups (e.g. Cooper et al., 2010; Farooq et al., 2021). This highlights the need for further research to address wider systemic factors which may contribute to vulnerable individuals seeking support online (De Riggi, Lewis, & Heath, 2018) and/or via friends and family (Fortune, Sinclair, & Hawton, 2008), such as difficulty in accessing mental health services and perceived stigma (e.g. negative professional responses to self-harm; Rowe et al., 2014; Saunders, Hawton, Fortune, & Farrell, 2012), or racism and poverty (Castro-Ramirez et al., 2021). Offering more timely and innovative support online and offline, such as having peer support mentors, in the community and online, could improve effective help-seeking. Another way this could be achieved is through greater dissemination of information and guidance being accessible online and via social media platforms to young people, parents/carers and any other individuals responsible for supporting young people (e.g. primary care, community and cultural groups, and schools). Schools in particular are important places to provide psychoeducation about online safety, facilitate discussion and identify those at risk. The #chatsafe project in Australia has developed guidelines to support young people to talk online safely

about suicide (Robinson et al., 2018). The findings presented here could be used to expand on such guidance, relating to viewing, sharing and commenting on self-harm-related images more specifically, such as the potential consequences for the self and others, in order to enable the protective mechanism of social support to continue, but at the same time, reducing the potential for harm. It would also be important to evaluate the impact of such guidance in reducing harmful impacts.

Policy implications

The findings of this review have implications for national policies for regulation of online communities, such as the UK's Online Safety Bill (Department for Digital, Culture Media and Sport, 2021). For example, the areas of potential impact identified could help guide the Bill's definitions of harm and need for regulation. The impact of recent changes, such as blocking self-harm hashtags or images, is unknown. More research into the effectiveness of trigger warnings is required, as some individuals may view images regardless of warnings (e.g. Baker & Lewis, 2013). Whilst images may require more restrictive regulation than verbal content, policy changes should be made following consultation with Experts by Experience and introduced alongside timely evaluation of the impact of changes.

Conclusions

This review explored the potential impacts and mechanisms of viewing self-harm images online. With new platforms being continually developed, awareness of the harmful and protective effects is vital. All 15 studies reported evidence relating to harmful effects, such as escalating self-harm behaviours, emotional distress, and reinforcement of engagement behaviours (e.g., commenting, sharing images). There was also some evidence of protective effects, such as through social support. Although causality of impact has been assumed, this is currently unknown. Most studies were cross-sectional and relied on retrospective self-report and did not directly test hypotheses related to mechanisms. Further high-quality research is urgently needed to determine the factors that contribute to the inconsistencies in harmful and protective effects and the associated psychological mechanisms. The conceptual model we have developed provides a framework that may guide further investigations. Individuals may access online self-harm images for emotional regulation/dysregulation and social connection, and clinicians should routinely assess exposure to self-harm images and determine the function and impacts of accessing such content, to guide targeted interventions. This is important because of the idiosyncratic impacts that viewing self-harm images can have on individuals and how these may change over time, such as through repeated exposure or through changes in other vulnerability or contextual factors. The types of impacts identified in this review may also help with drafting national online safety policy guidelines and inform guidance more generally about the potential consequences of viewing and sharing online content related to self-harm.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1. Definitions and examples of potential mechanisms.
Table S2. Inclusion and exclusion criteria.
Table S3. CASP quality ratings.
Table S4. Detailed summary of studies.
Table S5. Harmful and protective impact and quality of studies.
Table S6. Studies reporting potentially harmful and protective areas of impact and associated mechanisms.

Acknowledgements

Oxford Health NHS Foundation Trust funded K.S. to undertake this research as part of her Doctorate in Clinical Psychology at the Oxford Institute for Clinical Psychology Training and Research. The authors thank Julia Hallam, Outreach Librarian at Oxford Health NHS Foundation Trust, for her help with the search strategy, and the rest of the library team for providing timely access to relevant papers. K.S. and K.H. conceptualised the review. K.S. and F.G.-F., under the guidance of K.H., A.S. and R.K.B., performed the systematic review. All authors contributed to the analysis of the findings. K.S. drafted the manuscript, with critical review and substantial input from A.S., K.H., and R.K.B. All authors read and approved the manuscript prior to submission. K.H. is a member of the National Suicide Prevention Strategy for England Advisory Group. All other authors declare no direct conflict of interest related to this study.

Ethical statement

Ethical approval was not required as data were not collected directly from human participants.

Correspondence

Keith Hawton, Centre for Suicide Research, Department of Psychiatry, University of Oxford, Oxford, UK; Email: keith.hawton@psych.ox.ac.uk
Karima Susi, Step 4 Psychological Therapies, Nottinghamshire Healthcare NHS Foundation Trust, The Bungalows, 74 Portland Street, Kirkby-in-Ashfield, Nottingham NG17 7AG, UK; Email: karima.susi@nottshc.nhs.uk

Key points

- All 15 studies identified harmful effects of viewing online self-harm images, such as escalating self-harm behaviours, emotional distress and reinforcement of engagement behaviours (e.g. commenting, sharing images).
- Nine studies reported protective effects of viewing online self-harm images, such as social connection and support.
- There has been little in-depth investigation or discussion regarding the potential mechanisms associated with viewing self-harm images online, with only two studies evaluating evidence for possible contagion influences.
- On balance, viewing self-harm images on social media and online is likely to contribute to more harmful than protective impacts.
- Our model provides a basis for further testing, clinical assessment and intervention, and considerations for online harms policy and other relevant guidelines.

References

Arendt, F., Scherr, S., & Romer, D. (2019). Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. *New Media and Society*, *21*, 2422–2442.

Baker, T.G., & Lewis, S.P. (2013). Responses to online photographs of non-suicidal self-injury: A thematic analysis. *Archives of Suicide Research*, *17*, 223–235.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Bell, J. (2014). Harmful or helpful? The role of the internet in self-harming and suicidal behaviour in young people. *Mental Health Review Journal*, *19*, 61–71.

Bentley, K.H., Nock, M.K., & Barlow, D.H. (2014). The four-function model of nonsuicidal self-injury: Key directions for future research. *Clinical Psychological Science*, *2*, 638–656.

Berkowitz, L. (1984). Some effects of thoughts on anti- and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological Bulletin*, *95*, 410–427.

Brennan, C., Saraiva, S., Mitchell, E., Melia, R., Campbell, L., King, N., & House, A. (2022). Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence. *Journal of Public Mental Health*, *21*, 57–69.

Brown, R.C., Fischer, T., Goldwich, A.D., Keller, F., Young, R., & Plener, P.L. (2018). #cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, *48*, 337–346.

Brown, R.C., Fischer, T., Goldwich, D.A., & Plener, P.L. (2020). "I just finally wanted to belong somewhere"—Qualitative analysis of experiences with posting pictures of self-injury on Instagram. *Frontiers in Psychiatry*, *11*, 1–8.

Castro-Ramirez, F., Al-Suwaidi, M., Garcia, P., Rankin, O., Ricard, J.R., & Nock, M.K. (2021). Racism and poverty are barriers to the treatment of youth mental health concerns. *Journal of Clinical Child and Adolescent Psychology*, *50*, 534–546.

Cha, C.B., Hooley, J.M., Deming, C.A., D'Angelo, E.J., Glenn, J.J., Teachman, B.A., & Nock, M.K. (2016). Examining potential iatrogenic effects of viewing suicide and self-injury stimuli. *Psychological Assessment*, *28*, 1510–1515.

Chen, R., Wang, Y., Liu, L., Lu, L., Wilson, A., Gong, S., … & Ou, J. (2021). A qualitative study of how self-harm starts and continues among Chinese adolescents. *British Journal of Psychiatry Open*, *7*, 1–7.

Cialdini, R.B., Reno, R.R., & Kallgren, C.A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026.

Cicchetti, D.V., & Allison, T. (1971). A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology*, *11*, 101–110.

Cooper, J., Murphy, E., Webb, R., Hawton, K., Bergen, H., Waters, K., & Kapur, N. (2010). Ethnic differences in self-harm, rates, characteristics and service provision: Three-city cohort study. *The British Journal of Psychiatry*, *197*, 212–218.

Critical Appraisal Skills Programme (CASP). (2018a). CASP cohort study checklist. Available from: https://casp-uk.net/casp-tools-checklists/ [last accessed 21 February 2021].

Critical Appraisal Skills Programme (CASP). (2018b). CASP qualitative studies checklist. Available from: https://casp-uk.net/casp-tools-checklists/ [last accessed 21 February 2021].

Daine, K., Hawton, K., Singaravelu, V., Stewart, A., Simkin, S., & Montgomery, P. (2013). The power of the web: A systematic review of studies of the influence of the internet on self-harm and suicide in young people. *PLoS One*, *8*, e77555.

Darlow, B., Fullen, B.M., Dean, S., Hurley, D.A., Baxter, G.D., & Dowell, A. (2012). The association between health care professional attitudes and beliefs and the attitudes and beliefs, clinical management, and outcomes of patients with low back pain: A systematic review. *European Journal of Pain*, *16*, 3–17.

De Riggi, M.E., Lewis, S.P., & Heath, N.L. (2018). Brief report: Nonsuicidal self-injury in adolescence: Turning to the internet for support. *Counselling Psychology Quarterly*, *31*, 397–405.

Department for Digital, Culture Media and Sport. (2021). Draft online safety bill. Available from: https://www.gov.uk/government/publications/draft-online-safety-bill [last accessed 1 June 2021].

Farooq, B., Clements, C., Hawton, K., Geulayov, G., Casey, D., Waters, K., … & Kapur, N. (2021). Self-harm in children and adolescents by ethnic group: An observational cohort study from the multicentre study of self-harm in England. *The Lancet Child and Adolescent Health*, *5*, 782–791.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Fleiss, J.L., Levin, B., & Paik, M.C. (2003). *Statistical Methods for Rates and Proportions*, 3rd ed. Hobroken, NJ: John Wiley & Sons Inc. https://doi.org/10.1002/0471445428

Fortune, S., Sinclair, J., & Hawton, K. (2008). Help-seeking before and after episodes of self-harm: A descriptive study in school pupils in England. *BMC Public Health*, *8*, 1–13.
Fu, K.-W., Cheng, Q., Wong, P.W.C., & Yip, P.S.F. (2013). Responses to a self-presented suicide attempt in social media. *Crisis*, *34*, 406–412.
Hasking, P., Whitlock, J., Voon, D., & Rose, A. (2017). A cognitive-emotional model of NSSI: Using emotion regulation and cognitive processes to explain why people self-injure. *Cognition and Emotion*, *31*, 1543–1556.
Haw, C., Hawton, K., Niedzwiedz, C., & Platt, S. (2013). Suicide clusters: A review of risk factors and mechanisms. *Suicide and Life-threatening Behavior*, *43*, 97–108.
Hawton, K., Bale, L., Brand, F., Townsend, E., Ness, J., Waters, K., … & Geulayov, G. (2020). Mortality in children and adolescents following presentation to hospital after non-fatal self-harm in the multicentre study of self-harm: A prospective observational cohort study. *The Lancet Child and Adolescent Health*, *4*, 111–120.
Hawton, K., Harriss, L., Hall, S., Simkin, S., Bale, E., & Bond, A. (2003). Deliberate self-harm in Oxford, 1990-2000: A time of change in patient characteristics. *Psychological Medicine*, *33*, 987–995.
Hawton, K., Hill, N.T.M., Gould, M., John, A., Lascelles, K., & Robinson, J. (2020). Clustering of suicides in children and adolescents. *The Lancet Child and Adolescent Health*, *4*, 58–67.
Hetrick, S.E., Subasinghe, A., Anglin, K., Hart, L., Morgan, A., & Robinson, J. (2020). Understanding the needs of young people who engage in self-harm: A qualitative investigation. *Frontiers in Psychology*, *10*, 1–10.
Holmes, E.A., Blackwell, S.E., Burnett Heyes, S., Renner, F., & Raes, F. (2016). Mental imagery in depression: Phenomenology, potential mechanisms, and treatment implications. *Annual Review of Clinical Psychology*, *12*, 249–280.
Holmes, E.A., & Mathews, A. (2010). Mental imagery in emotion and emotional disorders. *Clinical Psychology Review*, *30*, 349–362.
Hooley, J.M., & Franklin, J.C. (2018). Why do people hurt themselves? A new conceptual model of nonsuicidal self-injury. *Clinical Psychological Science*, *6*, 428–451.
Jacob, N., Evans, R., & Scourfield, J. (2017). The influence of online images on self-harm: A qualitative study of young people aged 16-24. *Journal of Adolescence*, *60*, 140–147.
Jaroszewski, A.C., Kleiman, E.M., Simone, P.K., & Nock, M.K. (2020). First-person stimuli: Improving the validity of stimuli in studies of suicide and related behaviors. *Psychological Assessment*, *32*, 663–676.
Jarvi, S., Jackson, B., Swenson, L., & Crawford, H. (2013). The impact of social contagion on non-suicidal self-injury: A review of the literature. *Archives of Suicide Research*, *17*, 1–19.
Joiner, T. (2003). Contagion of suicidal symptoms as a function of assortative relating and shared relationship stress in college roommates. *Journal of Adolescence*, *26*, 495–504.
Joiner, T. (2005). *Why people die by suicide*. Cambridge, MA, USA: Harvard University Press.
Kensinger, E.A., & Schacter, D.L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(2), 110–126.
Kidger, J., Heron, J., Lewis, G., Evans, J., & Gunnell, D. (2012). Adolescent self-harm and suicidal thoughts in the ALSPAC cohort: A self-report survey in England. *BioMed Central Psychiatry*, *12*, 1–12.
Klonsky, E.D. (2007). The functions of deliberate self-injury: A review of the evidence. *Clinical Psychology Review*, *27*, 226–239.
Laerd Statistics. (2015). Cohen's kappa using SPSS Statistics. *Statistical tutorials and software guides*. Available from: https://statistics.laerd.com/
Laerd Statistics. (2021). Weighted kappa SPSS statistics. *Statistical tutorials and software guides*. Available from: https://statistics.laerd.com/premium/spss/wk/weighted-kappa-in-spss-9.php [last accessed 21 February 2021].
Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
Lavis, A., & Winter, R. (2020). #online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal of Child Psychology and Psychiatry*, *61*, 842–854.
Lewis, S.P., Heath, N.L., Sornberger, M.J., & Arbuthnott, A.E. (2012). Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. *Journal of Adolescent Health*, *51*, 380–385.
Lewis, S.P., & Seko, Y. (2016). A double-edged sword: A review of benefits and risks of online nonsuicidal self-injury activities. *Journal of Clinical Psychology*, *72*, 249–262.
Marchant, A., Hawton, K., Burns, L., Stewart, A., & John, A. (2021). Impact of web-based sharing and viewing of self-harm-related videos and photographs on young people: Systematic review. *Journal of Medical Internet Research*, *23*, e18048.
Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., … & John, A. (2017). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. *PLoS One*, *12*, e0181722.
Mars, B., Heron, J., Biddle, L., Donovan, J.L., Holley, R., Piper, M., … & Gunnell, D. (2015). Exposure to, and searching for, information about suicide and self-harm on the internet: Prevalence and predictors in a population based cohort of young adults. *Journal of Affective Disorders*, *185*, 239–245.
Mitchell, K.J., & Ybarra, M.L. (2007). Online behavior of youth who engage in self-harm provides clues for preventive intervention. *Preventive Medicine*, *45*, 392–396.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *British Medical Journal*, *339*, 1–8.
Mueller, A.S., & Abrutyn, S. (2016). Adolescents under pressure: A new durkheimian framework for understanding adolescent suicide in a cohesive community. *American Sociological Review*, *81*, 877–899.
National Institute for Health and Care Excellence. (2022). *Self-harm: assessment, management and preventing recurrence* [NICE guideline No. 225]. Available from: https://www.nice.org.uk/guidance/NG225 [last accessed 23 September 2022].
Nock, M.K. (2009). Why do people hurt themselves? New insights into the nature and functions of self-injury. *Current Directions in Psychological Science*, *18*, 78–83.
Nock, M.K., & Prinstein, M.J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of Consulting and Clinical Psychology*, *72*, 885–890.
Padmanathan, P., Biddle, L., Carroll, R., Derges, J., Potokar, J., & Gunnell, D. (2018). Suicide and self-harm related internet use: A cross-sectional study and clinician focus groups. *Crisis*, *39*, 469–478.
Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., … & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC methods programme*. Available from: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [last accessed 23 May 2019].
Rimal, R.N., & Real, K. (2003). Understanding the influence of perceived norms on behaviors. *Communication Theory*, *13*, 184–203.

Robinson, J., Cox, G.R., Bailey, E., Hetrick, S.E., Rodrigues, M., Fisher, S., & Herrman, H. (2016). Social media and suicide prevention: A systematic review. *Early Intervention in Psychiatry*, *10*, 103–121.

Robinson, J., Hill, T.M., Thorn, P., Battersby, R., Theh, Z., Reavley, N.J., … & Skehan, J. (2018). The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: A Delphi study. *PLoS One*, *13*, e0206584.

Rodham, K., Hawton, K., & Evans, E. (2004). Reasons for deliberate self-harm: Comparison of self-poisoners and self-cutters in a community sample of adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*, 80–87.

Rodway, C., Tham, S.G., Richards, N., Ibrahim, S., Turnbull, P., Kapur, N., & Appleby, L. (2022). Online harms? Suicide-related online experience: A UK-wide caseseries study of young people who die bysuicide. *Psychological Medicine*, *1–12*, 1–12.

Rowe, S.L., French, R.S., Henderson, C., Ougrin, D., Slade, M., & Moran, P. (2014). Help-seeking behaviour and adolescent self-harm: A systematic review. *Australian and New Zealand Journal of Psychiatry*, *48*, 1083–1095.

Ryan-Vig, S., Gavin, J., & Rodham, K. (2019). The presentation of self-harm recovery: A thematic analysis of YouTube videos. *Deviant Behavior*, *40*, 1596–1608.

Saunders, K.E.A., Hawton, K., Fortune, S., & Farrell, S. (2012). Attitudes and knowledge of clinical staff regarding people who self-harm: A systematic review. *Journal of Affective Disorders*, *139*, 205–216.

Seko, Y., Kidd, S.A., Wiljer, D., & McKenzie, K.J. (2015). On the creative edge: Exploring motivations for creating non-suicidal self-injury content online. *Qualitative Health Research*, *25*, 1334–1346.

Seko, Y., & Lewis, S.P. (2018). The self—Harmed, visualized, and reblogged: Remaking of self-injury narratives on Tumblr. *New Media and Society*, *20*, 180–198.

Sternudd, H.T. (2012). Photographs of self-injury: Production and reception in a group of self-injurers. *Journal of Youth Studies*, *15*, 421–436.

Winkielman, P., & Gogolushko, Y. (2018). Influence of sub-optimally and optimally presented affective pictures and words on consumption-related behavior. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02261

World Health Organization. (2019). Suicide. Available from: https://www.who.int/news-room/fact-sheets/detail/suicide [last accessed 5 May 2020].

Accepted for publication: 25 November 2022