SPECIAL CONTRIBUTION

Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation

Simon Rice, PhD¹,²,³; Jo Robinson, PhD¹,²; Sarah Bendall, PhD¹,²; Sarah Hetrick DPsych¹,²; Georgina Cox PhD¹,²; Eleanor Bailey BA(Hons)¹,²; John Gleeson PhD⁴; Mario Alvarez-Jimenez PhD¹,²

Abstract

Objective: Suicide remains a major global public health issue for young people. The reach and accessibility of online and social media-based interventions herald a unique opportunity for suicide prevention. To date, the large body of research into suicide prevention has been undertaken atheoretically. This paper provides a rationale and theoretical framework (based on the interpersonal theory of suicide), and draws on our experiences of developing and testing online and social media-based interventions. Method: The implementation of three distinct online and social media-based intervention studies, undertaken with young people at risk of suicide, are discussed. We highlight the ways that these interventions can serve to bolster social connectedness in young people, and outline key aspects of intervention implementation and moderation. Results: Insights regarding the implementation of these studies include careful protocol development mindful of risk and ethical issues, establishment of suitably qualified teams to oversee development and delivery of the intervention, and utilisation of key aspects of human support (i.e., moderation) to encourage longer-term intervention engagement. Conclusions: Online and social media-based interventions provide an opportunity to enhance feelings of connectedness in young people, a key component of the interpersonal theory of suicide. Our experience has shown that such interventions can be feasibly and safely conducted with young people at risk of suicide. Further studies, with controlled designs, are required to demonstrate intervention efficacy.

Key Words: suicide prevention, youth mental health, early intervention, social media, online intervention

Résumé

Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation

The online environment provides distinct challenges and opportunities for suicide prevention and intervention for young people. While the application of online and social media-based interventions raise important practical, ethical and safety issues (Alvarez-Jimenez & Gleeson, 2012; Gleeson et al., 2014), once developed, the reach and relative low cost of these approaches can overcome existing service delivery gaps. Relative to the respective burden of disease, engagement with, and availability of mental health services fall disappointingly below those for physical ill health (Prince et al., 2007), and this is especially true for adolescents—a period known for the onset of highly prevalent and impairing mental disorders such as depression and anxiety (Patel, Flisher, Hetrick & McGorry, 2007). For those in the 10-19 age range, suicide is the leading cause of death globally (WHO, 2014) and innovative solutions are required to tackle this challenge. Given the individual, social and economic impacts associated with youth suicide, there is an imperative for the field to develop and disseminate accessible and engaging psychosocial interventions for young people at risk (Robinson & Pirkis, 2014).

Online and social media-based interventions

Due to their rapidly evolving nature, online and social media-based interventions are expected to become increasingly appealing to young people over the next decade and beyond. Internet use among young people is ubiquitous, with upwards of 95% of young people using the internet daily (Burns et al., 2013). Given their immediacy, 24-hour accessibility, and geographical scope, online and social media-based interventions have potential to reach young people who may not be inclined or able to seek help from traditional (i.e., face-to-face) sources. Recent systematic reviews have highlighted the potential effectiveness of online and social media-based interventions for young people at risk of suicide (Rice et al., 2014a; Robinson et al., 2015b). Reinforcing the significance of this emerging literature, as part of the 2013-2020 Mental Health Action Plan, the World Health Organization has set international benchmarks for e-mental health innovation, including expansion of responsive community-based services using electronic and mobile technologies, cohesive online professional and peer-support, stepped-care, and engaging self-help (WHO, 2013).

The development and dissemination of ‘next generation’ e-mental health interventions, incorporating social media-based approaches, is a necessary step towards broadening acceptable and accessible support for young people at risk of suicide. Social media is rapidly becoming an essential aspect of social communication, particularly among young people. It is likely that social media will become pivotal in young people’s engagement with mental health services in the future (Appelbaum & Kopelman, 2014). However, given that research on the use of social media-based interventions is only in its early stages, there are significant knowledge and practice gaps (Rice et al., 2014a; Robinson et al., 2015b), and available technology is not yet being fully harnessed by youth health services (Montague, Varcin, Simmons & Parker, 2015). While initial work to develop guidelines in this area has commenced (TeamUp, 2015), it will be some time before best-practice in the field is identified.

Below we outline a theoretical approach for how online and social media-based interventions may reduce suicide risk in young people, and discuss strategies regarding the implementation of such interventions. Written for clinicians and researchers working with adolescent populations, the paper provides a practical overview of key considerations related to protocol development, management of ethical review and approval, intervention safety, implementation of ongoing team-based moderation and identification of key areas of future enquiry. We highlight that social media-based interventions are an emerging and dynamic area of inquiry, with significant opportunity for user-friendly and targeted interventions, and it is our hope that this paper encourages other research groups to build on, and extend our work.

Suicide risk – An interpersonal perspective

To date, much of the empirical research work within the suicide field has been conducted from an atheoretical approach (Van Orden et al., 2010). This has resulted in an abundance of checklists of historical risk factors (with relatively low predictive ability), rather than a richer, real-time explanation of etiological processes underlying suicide (Wenzel & Beck, 2008). In predicting suicide risk in a cohort of children and adolescents with depression, age, depression severity, hopelessness and family functioning were all significant predictors (Hetrick, Parker, Robinson, Hall...
& Vance, 2012). However the precise nature of how such predictors operate to precipitate a suicidal event remains unclear (Brent, 2011; Wenzel & Beck, 2008).

According to the interpersonal theory of suicide (Joiner, 2007; Van Orden et al., 2010), those with a marked risk for suicide are likely to experience: (i) a perceived sense of burdensomeness; and, (ii) a marked sense of isolation and disconnection from others. In the event of a suicide attempt, these factors are theorised to interact with a learned ability to enact lethal self-harm (i.e., overcoming the human instinct for self-preservation). The interpersonal theory of suicide provides a framework for targeting modifiable risk factors, and is seen by many as offering major advantages beyond current epidemiological models (Christensen, Batterham, Soubelet & Mackinnon, 2013). Interventions delivered through online and social media platforms are uniquely placed to tackle these interpersonal factors. Specifically, we argue that online and social media-based interventions can address the human need to belong (i.e., reducing a sense of thwarted belongingness) while also potentially moderating feelings of perceived burdensomeness. Online and social media-based interventions have the capacity to do this by enhancing the likelihood of social connectedness and reducing social isolation, factors that are associated with suicidal ideation in young people (Tang & Qin, 2015). Online and social media-based interventions can also provide opportunities for help seeking individuals to provide peer support to others (thus bolstering feelings of competence and autonomy). In doing so, the right type of online and social media interventions can counteract social withdrawal, and provide much needed opportunities for the development of meaningful, positive and supportive relationships.

**Implementation of online and social media-based interventions for suicide risk**

We draw on our own experience to highlight the importance of a well-coordinated, team-based approach to safely supporting young people at risk of suicide participating in online and social media-based interventions. This is based on findings from a recent survey of organisations and
Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation

Critical first steps in the evaluation of online and social media-based interventions for suicide risk include the development of protocols and ethical approval. Our experience, and undoubtedly the experience of many other research groups, is that ethics committees approach projects related to suicide risk with a high degree of caution (Lakeman & FitzGerald, 2009). This is unfortunate, as due to the lack of well-designed intervention studies, the active ingredients of treatment for the reduction of suicidal ideation and behaviour remain largely unknown (Robinson & Pirkis, 2014). Given randomised controlled trials frequently exclude participants experiencing suicidal ideation (Hetrick, McKenzie, Cox, Simmons & Merry, 2012), it may not be appropriate to simply import empirically supported treatments for young people experiencing depression to young people experiencing suicidality (Stanley et al., 2009).

Obtaining ethical approval for online and social media-based interventions for young people at risk of suicide is assisted through the development of clear and comprehensive risk management protocols, incorporating contingencies for a broad range of potential situations, including very low probability events. This may include the need to establish an independent safety advisory committee to oversee the conduct of the study, and consideration of postvention strategies prior to implementation (i.e., how other members of an online community can be supported following a suicide). In managing the ethical approval process, it may helpful to draw on the evidence base that demonstrates asking young people about suicide does not increase risk of suicidality (rather such enquiry decreases suicide risk) (Mathias et al., 2012; Robinson et al., 2011). For intervention-based studies, a safety plan should be developed collaboratively between the researcher and the young person when they are about to begin the intervention, and this plan should remain accessible to the young person (ideally via the online platform being tested/used) and to other key support people throughout their participation, and ideally beyond.

Components of moderation

A team-based approach to the moderation of online and social media-based interventions provides opportunity for joint decision making and shared responsibility, and should be facilitated by an experienced clinician or researcher with sufficient supervisory experience and training (Rice et al., 2014b). Of note, the competence and expertise of moderation was viewed by experts in the field as a key concern when responding to young people at risk of suicide via social media-based platforms (Robinson et al., 2015c). Senior staff must ensure that relevant safety protocols are adhered to by team members. Regular, structured team meetings (i.e., with a consistent agenda, clear roles and responsibilities, lines of accountability) help to provide containment to the team, assist with managing clinician or researcher anxiety, and offer a framework for clinical decision making. There should be sufficient time dedicated in team meetings to safety planning, including discussion of the need to contact support people or caregivers where indicated, and ad hoc (i.e., phone-based) supervisory support should be available if needed. Participant confidentiality must be maintained within the duty of care framework (i.e., participants are made aware of limits to confidentiality if risk to self escalates).

Any interventions undertaken with young people at risk of suicide that enable posting of user-generated content must be regularly moderated (i.e., at least daily). Developing a moderation roster, with an allocated supervisor, is recommended. As young people are known to express their suicidal ideation on social networking sites, and occasionally use these media to find out about suicidal methods (Cheng, Kwok, Zhu, Guan & Yip, 2015), monitoring of online platforms should occur at times that allow any necessary follow-up. Follow-up should include clear strategies for managing escalation in participant risk, or posting of inappropriate content. For example, monitoring of a school-based suicide prevention intervention should occur with sufficient time to contact school wellbeing staff and develop a short-term plan, which may include parental contact or referral for more intensive mental health support (Robinson et al., 2015b).

Protocol development and approval

Critical first steps in the evaluation of online and social media-based interventions for suicide risk include the development of protocols and ethical approval. Our experience, and undoubtedly the experience of many other research groups, is that ethics committees approach projects related to suicide risk with a high degree of caution (Lakeman & FitzGerald, 2009). This is unfortunate, as due to the lack of well-designed intervention studies, the active ingredients of treatment for the reduction of suicidal ideation and behaviour remain largely unknown (Robinson & Pirkis, 2014). Given randomised controlled trials frequently exclude participants experiencing suicidal ideation (Hetrick, McKenzie, Cox, Simmons & Merry, 2012), it may not be appropriate to simply import empirically supported treatments for young people experiencing depression to young people experiencing suicidality (Stanley et al., 2009).

Obtaining ethical approval for online and social media-based interventions for young people at risk of suicide is assisted through the development of clear and comprehensive risk management protocols, incorporating contingencies for a broad range of potential situations, including very low probability events. This may include the need to establish an independent safety advisory committee to oversee the conduct of the study, and consideration of postvention strategies prior to implementation (i.e., how other members of an online community can be supported following a suicide). In managing the ethical approval process, it may helpful to draw on the evidence base that demonstrates asking young people about suicide does not increase risk of suicidality (rather such enquiry decreases suicide risk) (Mathias et al., 2012; Robinson et al., 2011). For intervention-based studies, a safety plan should be developed collaboratively between the researcher and the young person when they are about to begin the intervention, and this plan should remain accessible to the young person (ideally via the online platform being tested/used) and to other key support people throughout their participation, and ideally beyond.

Components of moderation

A team-based approach to the moderation of online and social media-based interventions provides opportunity for joint decision making and shared responsibility, and should be facilitated by an experienced clinician or researcher with sufficient supervisory experience and training (Rice et al., 2014b). Of note, the competence and expertise of moderation was viewed by experts in the field as a key concern when responding to young people at risk of suicide via social media-based platforms (Robinson et al., 2015c). Senior staff must ensure that relevant safety protocols are adhered to by team members. Regular, structured team meetings (i.e., with a consistent agenda, clear roles and responsibilities, lines of accountability) help to provide containment to the team, assist with managing clinician or researcher anxiety, and offer a framework for clinical decision making. There should be sufficient time dedicated in team meetings to safety planning, including discussion of the need to contact support people or caregivers where indicated, and ad hoc (i.e., phone-based) supervisory support should be available if needed. Participant confidentiality must be maintained within the duty of care framework (i.e., participants are made aware of limits to confidentiality if risk to self escalates).

Any interventions undertaken with young people at risk of suicide that enable posting of user-generated content must be regularly moderated (i.e., at least daily). Developing a moderation roster, with an allocated supervisor, is recommended. As young people are known to express their suicidal ideation on social networking sites, and occasionally use these media to find out about suicidal methods (Cheng, Kwok, Zhu, Guan & Yip, 2015), monitoring of online platforms should occur at times that allow any necessary follow-up. Follow-up should include clear strategies for managing escalation in participant risk, or posting of inappropriate content. For example, monitoring of a school-based suicide prevention intervention should occur with sufficient time to contact school wellbeing staff and develop a short-term plan, which may include parental contact or referral for more intensive mental health support (Robinson et al., 2015b).

Protocol development and approval

Critical first steps in the evaluation of online and social media-based interventions for suicide risk include the development of protocols and ethical approval. Our experience, and undoubtedly the experience of many other research groups, is that ethics committees approach projects related to suicide risk with a high degree of caution (Lakeman & FitzGerald, 2009). This is unfortunate, as due to the lack of well-designed intervention studies, the active ingredients of treatment for the reduction of suicidal ideation and behaviour remain largely unknown (Robinson & Pirkis, 2014). Given randomised controlled trials frequently exclude participants experiencing suicidal ideation (Hetrick, McKenzie, Cox, Simmons & Merry, 2012), it may not be appropriate to simply import empirically supported treatments for young people experiencing depression to young people experiencing suicidality (Stanley et al., 2009).

Obtaining ethical approval for online and social media-based interventions for young people at risk of suicide is assisted through the development of clear and comprehensive risk management protocols, incorporating contingencies for a broad range of potential situations, including very low probability events. This may include the need to establish an independent safety advisory committee to oversee the conduct of the study, and consideration of postvention strategies prior to implementation (i.e., how other members of an online community can be supported following a suicide). In managing the ethical approval process, it may helpful to draw on the evidence base that demonstrates asking young people about suicide does not increase risk of suicidality (rather such enquiry decreases suicide risk) (Mathias et al., 2012; Robinson et al., 2011). For intervention-based studies, a safety plan should be developed collaboratively between the researcher and the young person when they are about to begin the intervention, and this plan should remain accessible to the young person (ideally via the online platform being tested/used) and to other key support people throughout their participation, and ideally beyond.

Components of moderation

A team-based approach to the moderation of online and social media-based interventions provides opportunity for joint decision making and shared responsibility, and should be facilitated by an experienced clinician or researcher with sufficient supervisory experience and training (Rice et al., 2014b). Of note, the competence and expertise of moderation was viewed by experts in the field as a key concern when responding to young people at risk of suicide via social media-based platforms (Robinson et al., 2015c). Senior staff must ensure that relevant safety protocols are adhered to by team members. Regular, structured team meetings (i.e., with a consistent agenda, clear roles and responsibilities, lines of accountability) help to provide containment to the team, assist with managing clinician or researcher anxiety, and offer a framework for clinical decision making. There should be sufficient time dedicated in team meetings to safety planning, including discussion of the need to contact support people or caregivers where indicated, and ad hoc (i.e., phone-based) supervisory support should be available if needed. Participant confidentiality must be maintained within the duty of care framework (i.e., participants are made aware of limits to confidentiality if risk to self escalates).

Any interventions undertaken with young people at risk of suicide that enable posting of user-generated content must be regularly moderated (i.e., at least daily). Developing a moderation roster, with an allocated supervisor, is recommended. As young people are known to express their suicidal ideation on social networking sites, and occasionally use these media to find out about suicidal methods (Cheng, Kwok, Zhu, Guan & Yip, 2015), monitoring of online platforms should occur at times that allow any necessary follow-up. Follow-up should include clear strategies for managing escalation in participant risk, or posting of inappropriate content. For example, monitoring of a school-based suicide prevention intervention should occur with sufficient time to contact school wellbeing staff and develop a short-term plan, which may include parental contact or referral for more intensive mental health support (Robinson et al., 2015b).
et al., 2014c). A log of moderation activities should be kept in a secure, centralised location. This log should document actions taken and any actions pending. Any limits to moderation should be clearly indicated (i.e., days and times the platform is not moderated).

Moderators must be able to check all user-generated content posted to the social network or online, and respond appropriately if required. In complex systems, this may require software that displays all online activity within a given period, typically since last episode of moderation. In our experience, the vast majority of user-generated content within social media-based interventions is either positive in nature (i.e., posts reflecting on therapeutic content available within the intervention, posts offering emotional support to others) or benign (i.e., posts sharing a link to a favourite music video). In rare cases of user generated content that is ambiguous or negative (i.e., a post suggesting or clearly referring to a user’s current negative emotional state or clinical deterioration), direct contact should be made to immediately assess risk of harm to self. Where it is not possible to contact the user directly (either via email, text message or phone call) it may be necessary to phone the emergency contact person. Detailed notes on any such contact should be kept within a moderation log that is available to all moderators. In any such cases, the wider moderation team should be promptly notified of the management plan, which may include referral for face-to-face support, suggestions to complete specific therapeutic content available within the intervention (i.e., safety plan, content relevant to affect regulation) and/or closer online or direct-contact monitoring, until resolution of the episode (for more information see Alvarez-Jimenez et al. 2013; Robinson et al., 2014a; 2014b; 2014c).

Models highlighting the importance of human support in the delivery of online interventions have recently been developed. An example of this includes the moderated online social therapy (MOST) model, which underpins the Rebound intervention (Alvarez-Jimenez et al., 2013; Rice et al., in press). MOST aims to engage users over the longer term through a combination of expert clinician and peer moderation. Models such as MOST seek to enhance the social connectedness and clinical outcomes of users through the integration of moderator and self-directed evidence based therapy content and social networking. This occurs through an interactive newsfeed where users can post content and comment on other user generated material. Similarly, a human support moderation model was used in the social networking approach adopted in the Safe Conversations project (Robinson et al., 2015a). In this study a closed Facebook group enabled users to communicate both with each other, and the research team, and to provide real-time feedback on the interventions developed as part of this project.

A further essential component of the moderation of online and social media-based interventions involves attention to potential harms that may accompany platform use. Consideration should be given to cyberbullying, harassment and privacy concerns. Detailed strategies for managing these and other issues are listed elsewhere (Gleeson et al., 2014), but in closed platforms there may be a need to orientate users to conditions of use, outline prohibited behaviours (such as hostility or posting other inappropriate content) and ensure limitations on the availability of personally identifying information, including posting of user phone numbers or email addresses. Consideration should be given to managing the risk of ‘copycat’ events related to self-harm or suicide, and although rare, the concern that social networking sites may potentially be used for the formation of suicide pacts (Ozawa-de Silva, 2008). Such incidents may be managed by moderators (or automatically by the system) deleting or blocking distressing or provocative user-generated posts.

Maximising potential benefits

In line with the interpersonal theory of suicide, online and social networking interventions can be designed to enhance social connectedness and reduce perceived burdensomeness. For example, interacting with others with similar experiences may reduce feelings of isolation or separateness (Gilat & Shahar, 2009), promoting affiliative bonds and group belongingness. Further, online and social networking interventions also have the capacity to enable participants to offer peer support to others (Hsiung, 2007), which may have flow on effects of reducing perceived burdensomeness. Such reductions to social isolation may also be reinforcing, promoting longer term engagement with the intervention. The critical factor here is that online and social media-based interventions promote an authentic sense of belonging and meaningful relationships. This can be encouraged via a number of angles, including input from moderators, peer moderators, and the manner in which therapeutic content is provided, for example, as done via the MOST model. There may also be opportunities for the intervention to facilitate other forms of help seeking, such as accessing support from a school counsellor or mental health service (Robinson et al., 2014c).

There is significant scope for online and social media interventions to aid in the screening or identification of at-risk individuals. Given their immediate nature, these platforms can also be used to intervene in instances of emergency. New methodologies are developing that may enable the identification of suicide risk factors on a large (i.e., national) scale (Jashinsky et al., 2015). If used carefully, online and social media interventions have the capacity to target otherwise unreachable populations and broker suicide prevention messaging and interventions on a 24 hour basis (Robinson et al., 2015b). Immediate intervention may occur via the posting of status update posts or other user-generated posts.
content, or through the display of targeted messages (Moreno et al., 2012).

**Broader applications and future directions**

While the abovementioned studies recruited young people experiencing suicidal ideation, not all studies discussed (i.e., Rebound) included those with acute suicide risk. Indeed, the development of methodologies and protocols that focus on the high risk spectrum remains a key area for evaluation. Taking a broader view, there is significant scope for social media platforms to be used for suicide prevention (Robinson et al., 2015c). For example, the Safe Conversations project (Robinson et al., 2015a) provides a model for how this can be done safely within the context of Facebook. This study used the closed group functionality on Facebook to facilitate communication between researchers and participants in the development of a suite of suicide prevention interventions designed to be delivered via social media. This methodology also provide real-time user feedback on the interventions being developed, within a safe and supported environment. Moderated daily by a researcher, to date all communication via the Safe Conversations group has been appropriate and no adverse events have been reported. Further, as social networking companies become increasingly cognisant of the ways in which young people are using their services in relation to mental health help seeking, in the future there will likely be important collaborations between researchers, clinicians and technology companies.

While this paper outlines a broad rationale for youth suicide prevention within the online environment, we acknowledge that some of the ideas discussed are notional and require further testing, likely through new and emerging methodologies. As interventions for young people at risk of suicide begin to expand to true population-based approaches, moderation models will need to evolve to ensure a balance between user safety and usefulness (i.e., effectively helping those at risk). In addition, there is more work required in integrating content relevant for online and social media-based interventions, and whether social networking is truly indicative of a meaningful sense of connectedness. As we begin to discover the aspects of moderation that underpin a sense of connectedness in online and social media-based interventions, these factors can be maximised with the goal of reducing interpersonal risk factors associated with suicide (Joiner, 2009; Van Orden et al., 2010).

**Conclusions**

Our recent experience of providing online and social media-based interventions for young people at risk of suicide highlight the safety, acceptability and feasibility of these types of studies. We believe that such interventions may uniquely target, and reduce, important interpersonal risk factors for suicide. Our studies have shown that these types of interventions can be implemented in a safe manner. Controlled evaluations of these interventions are either underway, or planned, and will provide information on longer-term efficacy. It is acknowledged however that controlled studies are difficult to undertake using real world social media platforms. In the meantime, we encourage other groups to utilise, and test, online and social media-based platforms for supporting young people at risk of suicide. It is clear that young people seek out these platforms, and their use is all but certain to proliferate over the next generation.

**Acknowledgements/Conflicts of Interest**

This paper has been made possible from grants provided by the HCF Foundation, the National Health and Medical Research Council (NHMRC), Australian Catholic University, The Victorian State Government’s Mental Illness Research Fund, and the Telstra Foundation. SR has been supported via a Society for Mental Health Research Early Career Fellowship. JR has been supported by an Early Career Fellowship from the NHMRC. MAJ has been supported via a CR Roper Fellowship, Faculty of Medicine, Dentistry, and Health Science, The University of Melbourne, and a Career Development Fellowship (APP1082934) from the NHMRC.

**References**


