Dougall, Alyssa (2025) *The application of the Delphi methodology in intervention development for social withdrawal and Hikikomori.*
D Clin Psy thesis.

https://theses.gla.ac.uk/85437/

The Application of the Delphi Methodology in Intervention Development for Social Withdrawal and Hikikomori

Alyssa Dougall, BSc (Hons)

Submitted in partial fulfilment of the requirements for the degree of
Doctorate in Clinical Psychology

School of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University of Glasgow

February 2025

Table of Contents

List of Appendices

List of Tables

CHAPTER ONE: SYSTEMATIC REVIEW

CHAPTER TWO: MAJOR RESEARCH PROJECT

List of Figures

CHAPTER ONE: SYSTEMATIC REVIEW

CHAPTER TWO: MAJOR RESEARCH PROJECT

Chapter 1

The Use of Delphi Methodology in the Development of Psychologically Informed Interventions for Mental Health Conditions where Social Withdrawal is a Feature: A Systematic Review of the Literature

Prepared in accordance with the author requirements for Transcultural Psychiatry

Online Link to Submission Guidelines

Abstract

Background: Involving stakeholders in intervention development is part of best practice guidance to increase treatment efficacy and accessibility. The Delphi method allows for the systematic gathering of stakeholder opinions, though the reliability of these studies' conclusions depends on reporting quality.

Aim: To examine the application and reporting quality of studies using the Delphi method for developing psychologically informed interventions for mental health conditions where social withdrawal features.

Method: A systematic review was undertaken in September 2024 using the databases EMBASE. MEDLINE, PsychINFO, Psychology & Behavioural Sciences Collection and CINHAL. English language articles were included that described the use of the Delphi method for any aspect of intervention development for mental health conditions associated with social withdrawal. Data extraction contributed to a narrative synthesis of the methodology reporting quality, alongside a general quality appraisal of the study.

Results: Thirteen papers (2008-2024) were considered for data extraction. The evaluation found considerable variation in subject matter expert inclusion, preparation of surveys, and definition of consensus.

Conclusion: Despite the Delphi method being an appropriate approach for intervention development, there are significant discrepancies in the way that it is applied. This review demonstrated that researchers would benefit from referencing reporting guidance, to increase the transparency of their findings. While this review used a general guide for consensus studies, researchers may wish to use a method and speciality specific guide to enhance specificity of results. Due to the nature of social withdrawal, further efforts and adaptations may be required to increase stakeholder involvement.

Funding: *No funding was required for this review.*
Registration: *The review protocol was registered on Prospero (ID: CRD42024591853).*
Statements and Declarations: *The author(s) have no completing interests to declare.*
Key words: *Systematic Review; Delphi; Hikikomori; Social Withdrawal; Schizophrenia; Autism Spectrum Disorder; Social Anxiety; Avoidant Personality Disorder; Depression*

Introduction

Consensus-based Techniques in Intervention Development

Psychologically-informed interventions have become a first line treatment for common mental health disorders such as depression and anxiety (Cuijpers et al., 2023; Szuhany & Simon, 2022). However, with the prevalence and economic costs of mental health difficulties on the rise, clinical psychology is faced with important questions about how to enhance, or rebuild, current interventions (Rief et al., 2024). For complex mental health difficulties, researchers are recognising the benefit of extending the expertise involved in research (O'Cathain et al., 2019). The Medical Research Council's complex interventions framework encourages stakeholder involvement in intervention development: *"those individuals who are targeted by the intervention (or policy), those involved in its development or delivery, or those whose personal or professional interests are affected"* (Skivington et al., 2021; Craig et al., 2008).

The role of stakeholders can extend from consultation to coproduction, where decisions can be made regarding accessibility, content, and delivery of interventions (O'Cathain et al., 2019). Particularly in novel areas, or fields of uncertainty, stakeholder knowledge can build on scarce intervention evidence bases and enhance clinician understanding (King et al., 2021; Skivington et al., 2021). However, assembling information from a wide range of sources can lead to contradictions and non-systematic gathering of evidence (Sinha, Smyth & Williamson, 2011). As such, techniques have been developed to methodically gather diverse perspectives and meaningfully collate this to reach a shared understanding, or '*consensus*'. Gathering opinions in such a manner has led to more reliable output than unstructured group opinion (Kurvers et al., 2016). However, regardless of the type of consensus-based approach, the reliability of these conclusions depends on the transparency and quality of the application and reporting of consensus.

The Delphi Method

The Delphi technique is a consensus-based method designed to systematically collate expert opinions through iterative survey rounds (Jorm, 2015; Linstone & Turoff, 1975). The Delphi method was established in the 1950s for military purposes, and then developed for technological forecasting (Dalkey & Helmer-Hirschberg, 1963; Gordon & Helmer, 1964). Throughout time, the technique has been refined and continues to assist with complex problems in fields such as psychology, political science and sociology (Sprenkle & Piercy, 2005).

The Delphi method involves iterative rounds of consensus refinement with participants defined as ‘experts’ in the chosen area. Most commonly, the surveys request opinions as quantitative data (e.g. Likert scale ratings), however, qualitative information can be captured to inform or adapt survey items. Regardless of the nature of the data, a key aspect of the traditional Delphi methodology includes anonymity between participant ratings to encourage honesty, minimise groupthink and reduce social desirability bias (Sprenkle & Piercy, 2005; Habibi, Jahantigh & Sarafrazi, 2014).

Following the initial survey round, the levels of consensus are calculated as per the researchers’ chosen method (e.g. medians, interquartile ranges, percentage agreement; Sprenkle & Piercy, 2005) and compared against pre-defined consensus criteria. Dependent on this, items may be approved, excluded, adapted, or submitted for re-rating. The Delphi method uses several survey rounds and relies on feedback mechanisms that allow for the refinement of ideas, for example, sharing group means with subject matter experts (SMEs). This iterative process continues until pre-defined stopping criteria have been met (e.g. a set number of rounds (usually three) or stability statistics). The product of this method is a ‘consensus statement’ that includes the items that did and did not meet the consensus criteria.

The Delphi methodology can be used within many aspects of mental health, for example: developing outcome measures (Santaguida et al., 2018; Sinha et al., 2011); or treatment adaptation (King et al., 2021). However, this review will consider the role of the Delphi methodology in the development of psychologically-informed interventions. The systematic assembly of experts’ opinions can answer key uncertainties and ascertain priorities for all aspects of an intervention in an unbiased manner (e.g., determining which psychological processes to target; how meaningful change could be measured; Sinha et al., 2011).

The Delphi Method and Extreme Social Withdrawal

For more complex problems, particularly with significant individual and cultural dimensions, the Delphi method can be a useful starting point to intervention development (King et al., 2021). Extreme social withdrawal is a complex presentation, with the phenotype driven by a variety of mechanisms. For example, withdrawal can be underpinned by culturally-based factors (e.g. Hikikomori) or by mental health conditions with a range of complexity and duration (e.g. social anxiety, psychosis). Due to the variety of difficulties associated extreme social withdrawal, the treatment base for the phenotype is limited (Kato et al., 2020). Therefore, this systematic review will collate and critically analyse studies that use the Delphi methodology to develop psychologically-informed interventions for conditions characterised by extreme social withdrawal. A recent study established five categories of conditions that are associated with extreme social withdrawal and, as such, were selected to structure the focus of this systematic review. These categories include: schizophrenia and other psychotic disorders; social anxiety disorder; depression and associated conditions; autism spectrum disorder; avoidant personality disorder (Muris & Ollendick, 2023).

Rationale for the Review

Hikikomori and extreme social withdrawal are complex and multifarious phenotypes that may benefit from the use of a consensus-based approach to intervention development. More specifically, the evidence-base for existing syndrome-driven interventions is flawed, and so examining the application of the Delphi method in related conditions, may provide information regarding the usefulness of this technique (Rief et al., 2024).

The application of the Delphi method must be sufficiently rigorous for the output to be considered valid, however, the technique is not applied uniformly across studies (Sinha et al., 2011). There are often ambiguities about panel selection and size, stopping criterion and other aspects of the technique (Habibi et al., 2014). Attempts to operationalise quality criteria for the method have exposed variations in the application and reporting of existing studies (Nasa, Jain & Juneja, 2021; Diamond et al., 2014; Boulkedid et al., 2011; Jünger et al., 2017; von der Gracht, 2012). Additionally, such guidelines are recent developments and require ongoing validation (Gattrel et al., 2024; Sinha et al., 2011). Increased familiarity with quality appraisal tools and application guidance, may increase the use of this technique in other fields of great complexity and uncertainty, and support the recommended involvement of SMEs in a systematic manner.

Objectives

There is currently no comprehensive synthesis of studies that apply the Delphi method to psychologically informed intervention development. Given the lack of effective interventions for extreme social withdrawal, we will use this problem as a lens through which to examine the use of the consensus-based methods to develop new psychological interventions. As such, this review aims to consider:

1. *How often is the Delphi method used in the development of psychologically informed interventions for mental health conditions where social withdrawal is a feature?*
2. *For the included studies, what aspect of intervention development is the Delphi method used for (e.g. identifying priority treatment elements, establishing intervention content)?*
3. *For the included studies, how compliant is the Delphi methodology with the reporting standards outlined in the ACCORD framework?*

Methods

Protocol and Registration

The Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines were used for this review (Page et al., 2021; see Appendix 1.1 and 1.2). The protocol for this systematic review was registered on Prospero in September 2024 (ID: CRD42024591853).

Search Strategy

The literature search was conducted on 17/09/2024 on the following databases: EMBASE; MEDLINE; PsychINFO, Psychology & Behavioural Sciences Collection; and CINHAL. For each database, a search strategy was developed with support from specialist medical librarians (see Appendix 1.3). The sensitivity of the searches was piloted against five previously identified relevant studies. Additionally, a forward and backward citation search was conducted to check for the presence of additional eligible studies not identified by the electronic search.

Study Selection

Following electronic and manual de-duplication of the electronic search, all records were screened by title and abstract by the lead researcher based on pre-defined eligibility criteria (see Table 1.1). An independent rater reviewed a randomly selected 10% of the search retrieval, meeting 92% agreement of decisions. Remaining conflicts were discussed, and criteria was refined. Studies that may have met, or definitely met, inclusion criteria were then screened in full. The independent rater than reviewed 20% of the full-text screenings, meeting 100% agreement with the lead researcher.

Table 1.1: Inclusion and exclusion criteria for exported studies

Criterion	**Inclusion Criteria**	**Exclusion Criteria**
Study Design	Delphi methodology.	Non-Delphi studies or reviews/commentaries without original Delphi methodology

Population	Human participants (all relevant stakeholders, e.g. mental health professionals, researchers, experts by experience). Any age group.	Non-human participants
Interventions	Mental health-related interventions (psychosocial, behavioural) or interventions incorporating consideration of psychological processes	Interventions unrelated to mental health (e.g. physical health interventions without a mental health component or the application of the Delphi method in other fields).
Mental Health Categories	Focuses on mental health categories with known associations to extreme social withdrawal (Muris & Ollendick, 2023).	Mental health categories that are not cited by Muris & Ollendick (2023) as associated with extreme social withdrawal, or a focus on physical health domains.
Outcomes	Report outcomes of the Delphi process within an intervention development context (e.g. initial design, consensus on intervention components, piloting phase)	Unclear or incomplete reporting of outcomes or unrelated to intervention development (e.g. evaluation of previously existing treatment, development of outcome tools).
Publication type	English language. Empirical study. All publication types (peer reviewed articles, non-peer reviewed articles, dissertations, theses). Full text must be available.	Non-English languages. Posters or conference abstracts.
Publication date	Any time period.	No restrictions

Data Extraction

Data extraction was conducted by the lead researcher. In addition to including study characteristics, a standardised data extraction tool was developed from the 'ACcurate COnsensus Reporting Document' (ACCORD; Gattrel et al., 2024; see Appendix 1.4). The ACCORD framework is an EQUATOR (Enhancing the QUAlity and Transparency Of health Research) listed tool, and as such fits with the purpose of improving the quality of publications in health sciences via clear, evidence-based guidelines. This measure determines if any aspects of the reporting quality of the consensus approach may have impacted the reliability of the outcome (Gattrel et al., 2024). While the checklist includes good reporting practice for each phase of a research paper, this systematic review will focus on the 21-items relating to methodology only. The data extraction tool was piloted on five random papers to ensure applicability and appropriate use of the tool.

Synthesis of Results

Given the diversity of population, study designs, and interventions, a quantitative synthesis was not feasible and therefore this review was informed by a narrative synthesis framework (Popay et al., 2006). A preliminary synthesis of the data was based on the study characteristics and the domains outlined in the methodological section of the ACCORD framework (Gattrel et al., 2024). General themes and trends were then investigated, before analysing the general quality of the studies and the robustness of the systematic review strategy and reporting.

Quality Appraisal

A quality appraisal tool was derived from work by Nasa and colleagues (2021) who described nine quality indicators. These indicators assess the overall quality of the application of the Delphi method, as opposed to the quality of reporting only. Within this tool, nine components of the Delphi methodology can be rated (i.e. yes/no/unclear). This measure also acts as a second review of the papers (i.e. to ensure data extracted for the ACCORD framework is accurate) and to monitor the systematic review quality.

Results

Study Selection

2141 papers were exported into EndNote X9.0. A total of 1027 duplicates were initially removed, leaving 1114 papers to be screened by title and abstract. Following this, 1084 were excluded, leaving 30 papers to screen the full text. From this, 12 studies were eligible to include in the review. In addition to electronic searches, forward and backwards citation searching identified one further study (see Figure 1.1).

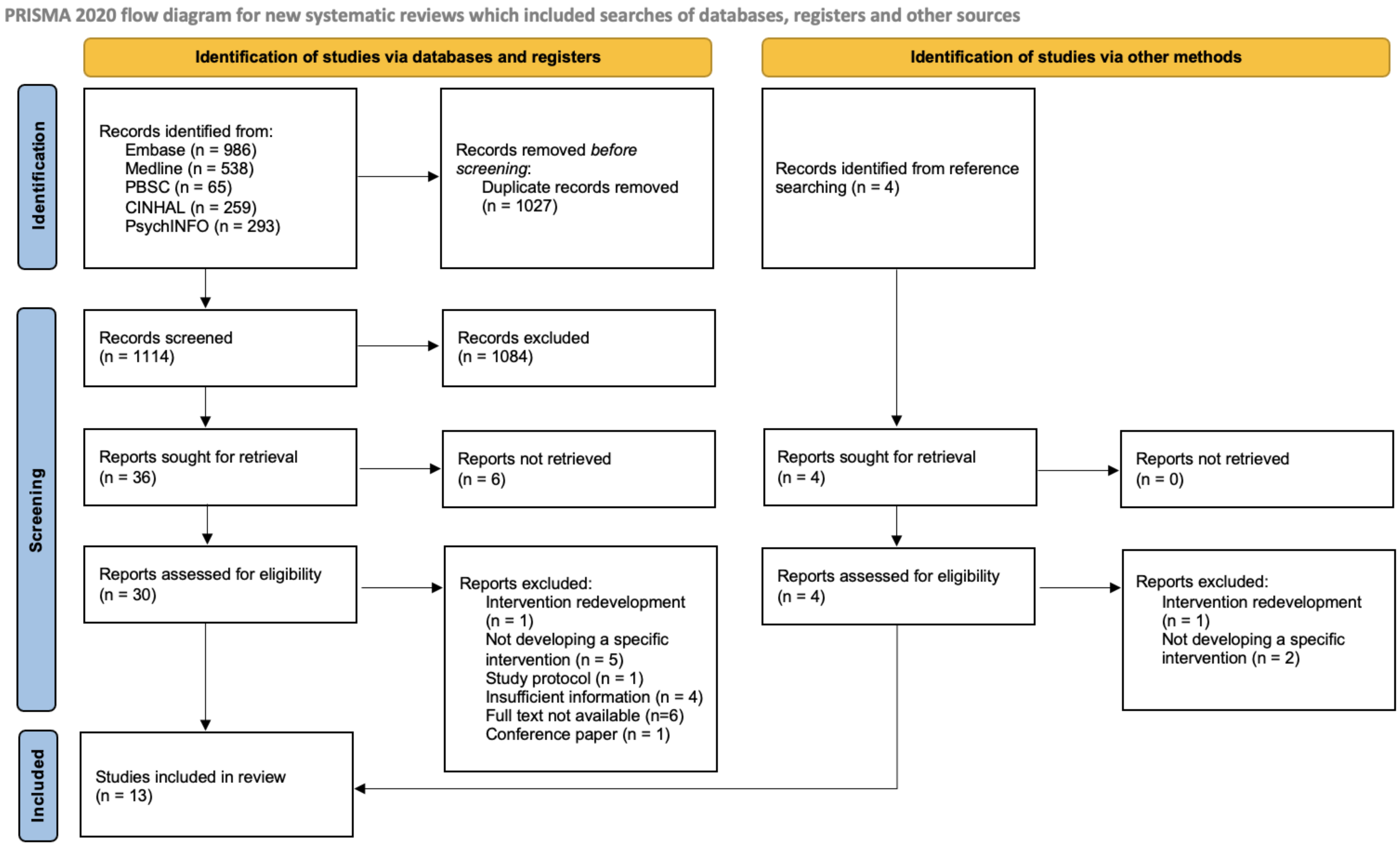

Figure 1.1: Prisma flow diagram for new systematic reviews which included searches of databases, registers and other sources. Source: Page MJ, et al. BMJ 2021;372: n71. Doi: 10.1136/bmj. Xn71.

Study Characteristics

A total of 13 articles from the year 2008-2024 were eligible (Table 1.2). Country of origin spanned Canada [*paper ID*: 1, 2], Brazil [3], Malaysia [4], United Kingdom [5, 8, 12], Australia [6, 9, 10, 11, 13] and Portugal [7].

Focus and purpose of the studies

All studies selected focused on developing a psychologically-informed intervention for a mental health condition associated with extreme social withdrawal (Muris & Ollendick, 2023). Five of the 13 studies focused on interventions relevant to neurodevelopmental and communication disorders [1, 2, 4, 6, 7], the majority of which were relevant to autism spectrum disorder specifically [1, 4, 6]. Four studies focused on mood disorders [3, 9, 10, 11]. One of these studies developed psychoeducational guidelines for bipolar [10], whereas the other three focused on depressive symptoms (sub-threshold and clinical; [3, 9, 11]). Three studies looked at psychosis [5, 12, 13], and one study focused on personality disorder [8].

All studies stated the aim to use the Delphi methodology to assist in intervention development and the majority aimed to reach consensus on the content and/or components of a future treatment [1, 2, 4, 5, 6, 8, 12]. Three studies focused on developing self-help and carer guidelines for those affected directly or indirectly by the target condition [9, 10, 11, 13]. Two used the Delphi method to develop an intervention through to the piloting stage [3, 7]. The specific nature of the interventions in development can be seen in Table 1.2.

Study design and type of Delphi

The majority of studies conducted their Delphi methodology online, with nine using an online or 'e-Delphi' approach [2, 3, 5, 6, 8, 9, 10, 11, 13], and two using an online 'modified Delphi' [1, 7]. While the specific modifications were unclear for one [1], one paper chose to not remove survey items or keep participants anonymous [7]. One study used a hybrid approach to their Delphi method, using online and in person surveys [12]. One study used a single round "fuzzy" Delphi method [[1]4].

---

[1] Fuzzy Delphi: A Delphi method that adjusts for uncertainty or inaccurate opinions by including 'fuzzy' data sets (i.e. ranges of responses rather than exact values; Murray, Pipino, van Gigch, 1985)

Table 1.2: The characteristics of the chosen studies including: the assigned paper ID; the primary author; the publishing year; the geographical area the study was based in; the mental health domain; the study aim; the intervention type; and the primary outcome(s).

ID	AUTHOR	YEAR	GEOGRAPH-ICAL AREA	MENTAL HEALTH DOMAIN	STUDY AIM	NATURE OF THE INTERVENTION	STUDY OUTCOME
1	Ghanouni et al.	2019	Canada	Children with autism spectrum disorders	To obtain consensus on the content of socio-emotional stories that can be used for a virtual reality programme.	Virtual reality programme	75 social story scenarios were validated that can be used to develop a virtual reality programme for children with autism spectrum disorder.
2	Ali et al.	2018	Canada	Children with neurodevelopmental disorders	To gain consensus about the key components of an eHealth parent-implemented intervention aimed at improving sleep in children with neurodevelopmental disorders.	Online, behavioural, parent-lead	52 priority items were identified to consider for an eHealth, parent-delivered intervention for insomnia in children with neurodevelopmental disorders.
3	Gonçalves et al.	2023	Brazil	Individuals with depressive symptoms	To develop a group intervention based on positive psychology to improve mental health outcomes in individuals with depressive symptoms.	12-session online group intervention	54/58 items met consensus, and the logistics and content of a 12-week online group intervention was validated and is now ready to pilot.

4	Ahmad Basri et al.	2024	Malaysia	"High functioning autism youth"	To determine the key components to design a social skills training content using virtual reality for youths with high functioning autism spectrum disorder.	Virtual reality social skills training	15/19 social skills difficulties and 11/15 cognitive and behaviour techniques met consensus criteria. These can be used as a framework for developing the training module in future studies.
5	Byrne & Morrison	2014	United Kingdom	Individuals with psychosis	To gain consensus about priorities and preferences for treatment of psychosis.	Therapeutic treatment priorities	17/64 domains of difficulty were rated as important/essential treatment priorities for individuals with psychosis.
6	Hodges et al.	2022	Australia	Elementary students on the Autism Spectrum	To gain consensus from experts on the content, delivery and feasibility of a school-based intervention that improves school participation and connectedness of elementary students with autism spectrum disorder.	Whole-class school-based intervention	Consensus on the intervention techniques was achieved for all relevant items.

7	Canossa Dias, Mineiro & Damen	2024	Portugal	Individuals with complex communication needs	To co-create and validate a new model and dynamic intervention program that analyses and improves communication between persons with "complex communication needs" and their caregivers.	A dynamic carer-based intervention programme including consultation, video analysis and psychoeducation.	All items were rated as adequate for an intervention that can be used by clinicians (e.g. speech and language therapists) in their work with caregivers of those with complex communication needs. The intervention can now be piloted.
8	Connell et al.	2022	United Kingdom	Justice-involved persons with personality disorder in the community	To gain expert consensus on an intervention to improve occupational participation including the components and their content; and to ascertain factors that influence on occupational participation, and the modifiability of these.	Intervention components.	121/148 statements met consensus agreement describing the content of an intervention to improve occupational participation. These statements will inform a further study that will develop a manualised intervention.

9	Langlands et al.	2008a	Australia	Individuals with sub-threshold depressive symptoms	To gain consensus on what elements mental health first aid for depression should involve.	Components for an early intervention, mental health first aid guidelines.	64/100 items met importance criteria when considering mental health first aid guidelines.
10	Berk et al.	2011	Australia	Caregivers of adults with bipolar	To develop guidelines to assist caregivers of adults with bipolar disorder.	Guidelines for caregivers	Roughly 537/626 (“85.8%”) items were endorsed. This resource can inform a stepped-care approach to supporting caregivers.
11	Cairns et al.	2015	Australia	Adolescents at risk of developing depression.	To ascertain elements of a self-help early intervention for depressive symptoms in adolescents.	Self-help preventative strategies.	145/194 items were endorsed and can now be promoted in developed English-speaking communities to help adolescents reduce their risk of depression.
12	Faulkner et al.	2022	United Kingdom	Individuals with schizophrenia and sleep difficulties.	Aimed to ascertain the contents and format of an occupational therapy intervention to improve sleep within the schizophrenia population	Components of an occupational therapy intervention for insomnia	The researchers did not report how many items consensus for importance, though said “many” of the survey items can be considered by clinicians

							when treating sleep problems in this population.
13	Langlands et al.	2008b	Australia	Individuals with psychosis	To gain consensus on recommended items for mental health first aid for psychosis.	Components for an early intervention, mental health first aid guidelines.	89/146 items met consensus for importance when considering mental health first aid guidelines.

Reporting Quality

The 21-items of the ACCORD framework that are relevant to methodological choices were used as reference for data extraction (see Appendix 1.4; Gattrel et al., 2024). For this review, the findings will be categorised into: registration; selection of panellists; preparatory research; assessing consensus; and participation (M20-M21). For each study, ACCORD items were rated as compliant (C), partially compliant (P), not compliant (N), or not applicable (N/A). Table 1.3 provides an overview of the 13 selected studies in line with the ACCORD methodological reporting guidance.

Registration

None of the papers explicitly stated within the study text that the research project had been preregistered.

Selection of steering committee and/or panellists

All 13 studies gave clear descriptions of the roles of experts and the study team who were directing the projects. Ten studies provided a clear explanation of panellist inclusion criteria, with three studies providing less detailed criteria [5. 9, 13]. All but one used professionals as participants (e.g. individuals with relevant qualifications, a set number of publications/years of experience, [1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13]). Four papers supplemented this information with input from parents or carers [1, 9, 10, 13] and/or patients/service users [9, 10, 11, 13]. One paper only used service-users as participants [5]. If stakeholders were involved in the study development, this was clearly noted [1, 3, 4, 5, 6, 7, 8 12] and took the form of survey item development and piloting the survey. For five studies it was unclear if any stakeholders played a role in the survey development process [2, 9, 10, 11, 13].

Additionally, seven studies described their rationale for sample size, with five studies referencing literature [3, 6, 7, 11, 12] and two demonstrating consideration of attrition [1, 8]. Six studies did not provide any justification for their chosen sample size [2, 4, 5, 9, 10, 13], though often this was reviewed as part of the discussion. Sample sizes ranged significantly from 10 [4] to 190 [10]. All 13 studies described the use of purposive sampling either on its own [2, 3, 4, 11], or in combination with another method such as convenience sampling [7, 9, 10, 12, 13] or snowball sampling [1, 8] or both [6].

Table 1.3: The overall ratings of the selected studies in line with the ACCORD framework's methodological items [2]*(Gattrel et al., 2024). For each item, ratings indicated compliance (C, green), partial compliance (P, orange), non-compliance (N, red), or were rated as not applicable (N/A, grey).*

Paper ID	Registration	Selection of SC and/or panellists				Preparatory research			Assessing consensus											Participation	
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21
1	N	C	P	C	C	C	N	N	C	C	C	C	C	C	P	P	P	N	P	N	P
2	N	C	P	C	C	C	P	N	C	C	C	C	C	C	C	N	C	C	C	C	N
3	N	C	C	C	C	C	P	N	P	C	C	P	P	C	P	P	N	N	C	N	P
4	N	C	P	C	C	C	P	N	C	C	C	C	N/A	C	C	C	P	N	C	N	P
5	N	C	P	C	C	C	N	N	C	C	C	P	C	C	C	P	N	N	P	N	P
6	N	C	C	C	C	C	N	N	C	C	C	C	N	C	C	C	C	C	P	C	C
7	N	C	C	C	C	C	N	C	C	C	C	C	N/A	C	C	C	C	C	C	C	C
8	N	C	P	C	C	C	P	N	C	C	C	C	C	C	P	C	P	N	P	C	C
9	N	C	P	C	C	C	C	N	C	C	C	P	C	C	P	N	P	N	P	N	N
10	N	C	P	C	C	C	P	C	C	C	C	C	C	C	P	N	C	C	P	C	C
11	N	C	C	C	C	C	C	C	C	C	C	C	C	C	P	N	P	N	P	N	N
12	N	C	C	C	C	C	N	C	P	C	C	P	P	C	P	P	P	N	P	C	P
13	N	C	P	C	P	C	C	N	C	C	C	P	C	C	P	N	P	N	P	N	N

---

[2] In line with the ACCORD framework, methodological quality items are labelled M1-M21. See Appendix 1.4 for an explanation of each item.

Preparatory Research

With regards to generating survey content, a literature review was most common [1, 2, 3, 4, 5, 6, 7, 9, 10, 11], with three papers describing more detailed search terms [3, 4, 10]. Three studies reported systematic review details within the study [9, 11, 13], and two cited a separate paper outlining the relevant systematic review [2, 8]. However, only four papers explicitly reported if, and how, the review findings were communicated to participants before or during the surveys [7, 10, 11, 12]. While six papers stated an information sheet was supplied to participants, they did not report evidence-based content of these and rather focused on participant role and ethical aspects of the project [2, 5, 6, 8, 9, 13].

As mentioned, stakeholder involvement was also used within survey development. In addition, some authors chose to pilot their survey(s) [4, 6, 7, 8]. However, these studies involving stakeholders required more detail on how many individuals were involved and in what form, the rationale for selecting these individuals, and the details regarding how their responses informed change. In four studies it was unclear if the steering committee or those managing consensus had voting rights in the surveys [2, 3, 4, 7].

Assessing Consensus

With regards to surveys, all 13 papers clearly explained how each survey item was presented, alongside the response options. All surveys used questions in quantitative format (e.g. Likert scale), however most also included open-field texts to gather qualitative data [1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13]. Five studies described how responses were processed and/or synthesised quantitively [4, 5] or using mixed-methods [2, 6, 7]. Seven of these studies provided unclear details of the qualitative analysis [1, 8, 9, 10, 11, 13] or the mixed-methods analysis [3, 12].

Various consensus definitions were used across studies. Ten provided a definition for consensus, and explained the rationale for this choice [1, 2, 3, 4, 6, 7, 8, 10, 11, 12] however, one of these studies did not explain their rules for item removal based on this [3], and another changed their interpretation throughout the study [12]. Nine studies chose to use percentage agreement as their only threshold for consensus, with thresholds including 70% [6], 75% [1, 8], 80% [3, 5, 9, 10, 11, 13]. Other studies relied on descriptive statistics such as means [2, 8], medians [7, 12], standard deviations [2], inter-quartile ranges [12] and interpercentile ranges adjusted for symmetry (IPRAS; [7]). One study used percentage agreement in addition to threshold values and fuzzy scores [4].

Reporting quality was variable for item removal following consensus criteria being met. Eight studies appropriately removed items based on pre-specified rules and criteria [1, 2, 5, 8, 9, 10, 11, 13]. Two reported removing items but the rules used were unclear [3, 12]. Two studies chose not to make removals based on study design [4, 7] and one paper did not report information regarding rules or removals [6].

The number of survey rounds varied from one [4], to two [1, 5, 6, 7], to three [2, 3, 8, 9,10, 11, 12, 13]. Over three quarters of studies provided feedback between rounds with three providing quantitative data in a clear manner [2, 7, 10] and one clearly described giving quantitative and qualitative feedback [6]. Seven studies would have been improved by providing more detail regarding the format of their quantitative [1, 8, 9, 11, 13] and mixed-method summaries [1, 12]. Two studies did not report any details of feedback being used, including the format of such [3, 5]. One study did not require feedback due to the nature of the design (fuzzy Delphi; [4]).

Participation

Most studies failed to explicitly confirm anonymity of their Delphi participants and feedback [1, 3, 4, 5, 8, 9, 11, 12, 13]. Only three studies reported anonymity [2, 6, 10] and one study explicitly reported the purposeful lack of anonymity [7].

Six studies reported attempts to encourage retention of participants either through financial incentives [8, 10, 12], and/or email reminders [2, 6, 7]. Four studies described specific adaptations to make the surveys more accessible such as altering the language [7], adjusting terminology and length [6, 8] and changing the modality of communication [7, 10]. Five other studies had previously used stakeholder involvement to contribute to survey development and so were partially complaint with the M21 criteria [1, 3, 4, 5, 12].

Quality Appraisal

All papers were reviewed in line with Nasa and colleagues' (2021) nine indicators for Delphi method quality (see Table 1.4). The ratings of this tool supported the data extracted for the

ACCORD tool indicating accuracy and continuity of application and reporting quality. Furthermore, 20% of the selected papers were blindly reviewed by the independent rater, meeting a weighted Cohens Kappa for inter-reliability statistic of 0.9, indicating ‘almost perfect agreement’ (Landis & Koch, 1977).

Table 1.4: A quality appraisal of the selected studies using the quality indicators outlined by Nasa and colleagues (2021). For each item, ratings indicated if studies met criteria (yes, green), did not (no, red), if it was unclear (unclear, orange) or if it was not applicable (N/A, grey).

ID	**1. Identifying problem area**	**2. Criteria for panellists**	**3. Anonymity**	**4. Controlled feedback**	**5. Iterative rounds**	**6. Consensus criteria**	**7. Consensus analysis**	**8. Closing criteria**	**Stability Rating**
1	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	No
2	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	No
3	Yes	Yes	Unclear	No	Yes	Yes	Yes	Yes	Unclear
4	Yes	Unclear	Unclear	N/A	N/A	Yes	Yes	N/A	N/A
5	Yes	Unclear	Unclear	No	Yes	Yes	*Yes*	Unclear	No
6	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
7	Yes	Yes	No	Yes	Yes	Yes	Yes	Unclear	No
8	Yes	Yes	*Unclear*	Yes	Yes	Yes	Yes	Yes	No
9	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Yes	*No*
10	Yes	Unclear	Yes	Yes	Yes	Yes	*Yes*	Unclear	No
11	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	No
12	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	No
13	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	Unclear	No

Discussion

This systematic review identified Delphi studies examining intervention development for conditions associated with extreme social withdrawal. Primarily, the method was used to ascertain the contents and/or components of the future intervention, indicating that the technique is most prominent in the information gathering, or priority identifying, stage of intervention development. The included studies had variable compliance with the reporting standards outlined in the ACCORD framework, demonstrating discrepancy in the transparency of the reporting of the technique.

Summary of Study Characteristics

Overall, 13 papers were reviewed. The studies ranged from 2008-2024 and there did not appear to be a particular trend in terms of the publication dates of relevant guidance (e.g. Delphi methodology or complex intervention development guidelines). The studies had wide geographical spread, with the most frequent area of publication being Australia. Four of the five Australian studies involved Anthony Jorm, who is a prominent researcher of the Delphi method in mental health contexts. As with other techniques, it is likely that influential protagonists increase the prominence of a method used in an area, and as such, this may have impacted the distribution of the results.

The most common administration method was an e-Delphi, with surveys being conducted online. One of two ‘modified Delphi’ studies elaborated on the specific adaptations made, and even with this, there were aspects of the data extraction tool that were no longer applicable. As the term “modified Delphi” is ambiguous, it is even more important that reporting guidance is followed and any changes should be justified (Hasson & Keeney, 2011; Keeney, Hasson & McKenna, 2006).

Reporting Quality of the Studies

The ACCORD framework was used to grade the reporting quality of the methodology of each paper. In general, studies explained the roles of those directing and supporting the study appropriately. In terms of participants, the recruitment process was well described, with particular prominence of purposive sampling methods. This is a common technique used in the Delphi method as panellist expertise corresponds to high quality outcomes (Dalkey, 1969). The pre-defined inclusion criteria were mostly clear, though the sample size was at times justified without literature backing. Required sample sizes will vary between studies dependent on resource availability, the homogeneity of the panel, and the complexity of the topic chosen (Nasa et al., 2021). However, consideration of attrition rate is key; as is reflected by four studies that reported high drop-out rates, particularly after the first survey. As attrition can lead the degree of consensus reached to be over-estimated, it is helpful to include strategies to minimise this, such as assessing interest, adapting accessibility of surveys, and using invitations to confirm interest in participating in advance of the study commencing (Bardecki, 1984; Sinha et al., 2011).

The studies primarily used professionals as participants, and as such the consensus may be excluding the priorities of those with lived experience (Sinha et al., 2011). It was difficult to assess the level of heterogeneity in the samples as, while the majority used professionals, these often represented various disciplines and qualifications. While heterogenous panels can mitigate cognitive biases and increase generalisation of the consensus, a homogenous panel may provide more reliable consensus in a particular study field, or a narrower focused study objective (Nasa et al., 2011; Beiderbeck et al., 2021). If using a heterogenous group, it can be beneficial to divide the participants into panels (as three studies did) or conduct an additional analysis on the distribution of ratings so that the opinions of minority groups are not masked (Goodman, 1987; Beiderbeck et al., 2021). This is particularly relevant for studies including those with socially withdrawn phenotypes, as they may be more likely to represent a smaller proportion of the participants, or to drop out.

With regards to study preparation, preregistration was not reported, which is recommended as best practice to reduce reporting bias (Kirkham et al., 2010). All studies did explain how survey items were developed in sufficient detail. This was primarily via a literature review, but other methods such as co-production with service users and interviews were also discussed. Piloting of surveys is also advised as it can provide insight into how the experts may interpret items and reduce accessibility concerns (which is particularly relevant for the socially withdrawn; Jünger et al., 2017). Collaborative involvement of stakeholders at this early stage is recommended to reduce the likelihood of researchers imposing their own views, which is why it is also recommended to provide qualitative opportunities within surveys. This minimises participants being confined by the literature and theory-driven survey items, encouraging experience-based and practically-led viewpoints; thus, impacting the final intervention (Sinha et al., 2011; Skivington et al., 2021).

The included studies indicate that the definition of ‘consensus’ in Delphi research requires further clarification and streamlining to be interpreted as meaningful. The description of consensus was reported in different ways across the studies in terms of the clarity of the definition, the type of consensus-method being chosen (e.g. percentage agreement), and the stability of the definition throughout the course of the study. Furthermore, the reporting of rules for item removal was poor, and as such it could be challenging to follow researcher decisions. However, all papers did report their consensus thresholds to some degree, which is an improvement from rates of 73% and 83% in previous reviews (Diamond et al., 2014; Jünger et al., 2017). Between rounds, feedback from experimenters to raters was poorly reported, which is highlighted as a concern given the significant impact of feedback on consensus (Scheibe, Skutsch & Schofer, 1975).

The Role of the Delphi Method in Intervention Development for Social Withdrawal

The Delphi method aligns with best practice guidelines for intervention development, such as the Medical Research Council (MRC) framework for complex interventions (Skivington et al., 2021). The MRC framework emphasises the meaningful contribution of stakeholders at each stage of intervention development in an iterative manner that mirrors the successive survey rounds in the Delphi method.

All studies selected for this review focused on developing psychologically informed interventions for mental health conditions associated with extreme social withdrawal (Muris & Ollendick, 2023). Due to the nature of social withdrawal, service engagement is typically poor, and so it is vital to encourage these individuals to share their own perspectives and priorities (Hareven et al., 2024). The anonymity of the Delphi method and the potential for online participation (as per most included studies) may assist in overcoming participation barriers. This is in line with the MRC framework’s emphasis on ascertaining stakeholder’s needs, challenges of engagement and flexibility of treatments (Craig et al., 2008; Skivington et al., 2021).

While some studies focused on the development of guidelines, or final refinements of an intervention pre-pilot, most studies aimed to assess consensus on the content or components of a future intervention. This highlights the role of the Delphi method in bottom-up approaches to intervention development, rather than top-down, protocol-driven treatments. Using stakeholders to collaboratively identify specific treatment components can allow for increased personalisation of treatments, which can be adaptive to culture or individual needs (as recommended in guidance; Skivington et al., 2021). This is particularly relevant for Hikikomori, where the underlying mechanisms contributing to the withdrawn phenotype, are understood to be highly variable and complex (Benarous et al., 2022; Muris & Ollendick, 2023).

Reporting Guidelines

While the included studies demonstrate that the Delphi method can be a beneficial approach, there is considerable discrepancy in the way that it is being used within intervention development. Such variety can affect the impact reliability of the results or the credibility of the recommendations (Gattrel et al., 2024). Additionally, in the absence of guidelines for good quality reporting, researcher perspectives and reporting bias can play a role and impact the meaning of the consensus (Sinha et al., 2011). The use of the ACCORD framework in this review was beneficial in capturing key methodological principles for the Delphi studies and is a suitable framework that can be followed by future researchers (Gattrel et al., 2024). There are other reporting guidelines that may be suitable for researchers conducting a Delphi method, however these often lack sufficient detail (e.g. Brouwers et al., 2011; Kirkham et al., 2016), or are speciality-specific (Jünger et al., 2017). Very recently, the DELPHISTARR (Delphi SelecTing Approaches for Rapid Reviews; Niederberger et al., 2024) was developed as a tool that aims to provide method-specific guidance to studies in the health and social sciences field. While this was published after this review had commenced, future studies may wish to draw on this guidance to increase the rigor and reliability of their studies.

Limitations of the Review

While the review was adherent to systematic review guidelines, and used an established tool for data extraction, the paper is not without limitations. As highlighted, the new

DELPHISTARR tool may be a more suitable framework for data extraction, as it is field and method-specific, though this was not developed at the time of designing the review (Niederberger et al., 2024). Additionally, while the screening and quality appraisal was supported by an independent reviewer, due to resource limitations, data extraction was conducted by one researcher. While the tool is well-structured and is accompanied with an explanation document, there is a chance of human error. Furthermore, the variability in reporting across studies, and the broadness of the mental health areas, limits the ability to draw conclusions on some aspects (e.g. do certain components of the methodology impact the final output?). Furthermore, while the review used a recent study to develop the inclusion criteria and conceptualise extreme social withdrawal, it must be noted that other frameworks may refer to additional relevant mental health conditions. Future researchers may wish to expand the scope of the review and refine the search strategies to reflect this and maximise the number of studies.

Conclusions

Complex intervention development guidelines for mental health conditions advocate for meaningful stakeholder involvement. Particularly in the context of developing treatment for conditions associated with extreme social withdrawal, the Delphi method can systematically assemble diverse perspectives when empirical data is scare. It’s anonymous, and often online, nature may encourage participation of those with lived experience of social withdrawal. This input may provide valuable insight into the barriers and incentives to treatment engagement.

While the Delphi method aligns with guidance for complex intervention development (e.g. iterative cycles), the application and reporting of the technique is known to be highly variable. As such, this review aimed to investigate this in the context of mental health conditions associated with extreme social withdrawal. The results indicated great inconsistency in the reporting quality of the Delphi method, particularly with regards to the rationale for recruiting certain SMEs and the sample sizes, the preparation of survey materials, and the definition of consensus.

As the reliability of the Delphi outcome is dependent on methodology and experts selected, reporting guidance should be used to increase the quality of the study and output. Future research should consider the use of a reporting quality checklist, such as the ACCORD framework, to increase the meaningfulness and impact of such stakeholder involvement in intervention development. Specifically, measures to include those with extreme social withdrawal in research should be an area of focus for future research.

References

Ahmad Basri, M. A. F., Wan Ismail, W. S., Kamal Nor, N., Mohd Tohit, N., Ahmad, M. N., Mohamad Aun, N. S., & Mohd Daud, T. I. (2024). Validation of key components in designing a social skills training content using virtual reality for high functioning autism youth-A Fuzzy Delphi method. *PloS one*, *19*(4), e0301517. https://doi.org/10.1371/journal.pone.0301517

Ali, N., Rigney, G., Weiss, S. K., Brown, C. A., Constantin, E., Godbout, R., Hanlon-Dearman, A., Ipsiroglu, O., Reid, G. J., Shea, S., Smith, I. M., Van der Loos, M., & Corkum, P. V. (2018). Optimizing an eHealth insomnia intervention for children with neurodevelopmental disorders: a Delphi study. *Sleep health*, *4*(2), 224–234. https://doi.org/10.1016/j.sleh.2017.12.008

Bardecki, M. J. (1984). Participants' response to the Delphi method: An attitudinal perspective. *Technological Forecasting and social change*, *25*(3), 281-292. https://doi.org/10.1016/0040-1625(84)90006-4

Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, *8*, 101401. https://doi.org/10.1016/j.mex.2021.101401

Benarous, X., Guedj, M. J., Cravero, C., Jakubowicz, B., Brunelle, J., Suzuki, K., & Cohen, D. (2022). Examining the hikikomori syndrome in a French sample of hospitalized adolescents with severe social withdrawal and school refusal behavior. *Transcultural psychiatry*, *59*(6), 831–843. https://doi.org/10.1177/13634615221111633

Berk, L., Jorm, A. F., Kelly, C. M., Dodd, S., & Berk, M. (2011). Development of guidelines for caregivers of people with bipolar disorder: a Delphi expert consensus study. *Bipolar disorders*, *13*(5-6), 556–570. https://doi.org/10.1111/j.1399-5618.2011.00942.x

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS one*, *6*(6), e20476. https://doi.org/10.1371/journal.pone.0020476

Brouwers, M. C., Kerkvliet, K., Spithoff, K., & AGREE Next Steps Consortium (2016). The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *British medical journal (clinical research ed.)*, *352*, i1152. https://doi.org/10.1136/bmj.i1152

Byrne, R., & Morrison, A. P. (2014). Service users' priorities and preferences for treatment of psychosis: a user-led Delphi study. *Psychiatric services (Washington, D.C.)*, *65*(9), 1167–1169. https://doi.org/10.1176/appi.ps.201300289

Cairns, K. E., Yap, M. B., Reavley, N. J., & Jorm, A. F. (2015). Identifying prevention strategies for adolescents to reduce their risk of depression: A Delphi consensus study. *Journal of affective disorders*, *183*, 229–238. https://doi.org/10.1016/j.jad.2015.05.019

Canossa Dias, J., Mineiro, A. & Damen, S. (2024). A new intervention to prevent social isolation in people with complex communication needs. *Scientific reports, 14*, 12742. https://doi.org/10.1038/s41598-024-63519-5

Connell, C., Furtado, V., McKay, E. A., & Singh, S. P. (2022). Developing an Intervention to Improve Occupational Participation for Justice-Involved People with a Personality Disorder: Defining and Describing Intervention Components. *International journal of offender therapy and comparative criminology*, *66*(6-7), 774–788. https://doi.org/10.1177/0306624X211013520

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., & Medical Research Council Guidance (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *British medical journal (clinical research ed.)*, *337*, a1655. https://doi.org/10.1136/bmj.a1655

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, *22*(1), 105–115. https://doi.org/10.1002/wps.21069

Dalkey, N. (1969). An experimental study of group opinion: The Delphi method. *Futures,* 1(5), 408-426. https://doi.org/10.1016/S0016-3287(69)80025-X

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, *9*(3), 458–467. Access: http://www.jstor.org/stable/2627117

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of clinical epidemiology*, *67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Faulkner, S. M., Drake, R. J., Ogden, M., Gardani, M., & Bee, P. E. (2022). A mixed methods expert opinion study on the optimal content and format for an occupational therapy intervention to improve sleep in schizophrenia spectrum disorders. *PloS one*, *17*(6), e0269453. https://doi.org/10.1371/journal.pone.0269453

Gattrel, W. T., Logullo, P., van Zuuren, E. J., Price, A., Hughes, E. L., Blazey, P., Winchester, C. C., Tovey, D., Goldman, K., Hungin, A. M., & Harrison, N. (2024). ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLOS Medicine*, https://doi.org/10.1371/journal.pmed.1004326

Ghanouni, P., Jarus, T., Zwicker, J. G., Lucyshyn, J., Mow, K., & Ledingham, A. (2019). Social Stories for Children with Autism Spectrum Disorder: Validating the Content of a Virtual Reality Program. *Journal of autism and developmental disorders*, *49*(2), 660–668. https://doi.org/10.1007/s10803-018-3737-0

Gonçalves, J. P., Braghetta, C. C., Alvarenga, W., Gorenstein, C., Lucchetit, G., & Vallada, H. (2023). Development of a comprehensive flourishing intervention to promote mental health using an e-Delphi technique. *Frontiers in Psychiatry, 14,* https://doi.org/10.3389/fpsyt.2023.1064137

Goodman C, M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing, 12,* 729-734. https://doi.org/10.1111/j.1365-2648.1987.tb01376.x

Gordon, T. J., & Helmer-Hirschberg, O. (1964). Report on a Long-Range Forecasting Study. Santa Monica, CA: RAND Corporation. Access: https://www.rand.org/pubs/papers/P2982.html.

Habibi, A., Jahantigh, F. F., & Sarafrazi, A. (2015). Fuzzy Delphi technique for forecasting and screening items. *Asian journal of research in business economics and management, 5,* 130-143. DOI: 10.5958/2249-7307.2015.00036.5

Hareven, O., Kron, T., Roe, D., & Koren, D. (2024). First-hand reports of prolonged social withdrawal: contributing factors, experiences, and change processes. *Mental Health and Social Inclusion*, *28*(5), 637-647. https://doi.org/10.1108/MHSI-05-2023-0055

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological forecasting and social change, 78*(9), 1695–1704. https://doi.org/10.1016/j.techfore.2011.04.005

Hodges, A., Cordier, R., Joosten, A., & Bourke-Taylor, H. (2022). Expert consensus on the development of a school-based intervention to improve the school participation and connectedness of elementary students on the autism spectrum: A Delphi study. *Focus on Autism and Other Developmental Disabilities, 37*(1), 13-23. https://doi.org/10.1177/10883576211030483

Jorm A. F. (2015). Using the Delphi expert consensus method in mental health research. *The Australian and New Zealand Journal of Psychiatry, 49*(10), 887–897. https://doi.org/10.1177/0004867415600891

Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative medicine*, *31*(8), 684–706. https://doi.org/10.1177/0269216317690685

Kato, T. A., Kanba, S., & Teo, A. R. (2020). Defining pathological social withdrawal: proposed diagnostic criteria for hikikomori. *World Psychiatry, 19*(1), 116-117. https://doi.org/10.1002/wps.20705

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing, 53*(2), 205–212. https://doi.org/10.1111/j.1365-2648.2006.03716.x

King, P. R., Jr., Beehler, G. P., Donnelly, K., Funderburk, J. S., & Wray, L. O. (2021). A practical guide to applying the Delphi technique in mental health treatment adaptation: The example of Enhanced Problem-Solving Training (E-PST). *Professional psychology: research and practice, 52*(4), 376– 386. https://doi.org/10.1037/pro0000371

Kirkham, J. J., Gorst, S., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., Moher, D., Schmitt, J., Tugwell, P., Tunis, S., & Williamson, P. R. (2016). Core Outcome Set-STAndards for Reporting: The COS-STAR Statement. *PLoS medicine*, *13*(10), e1002148. https://doi.org/10.1371/journal.pmed.1002148

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., Argenziano, G., Zalaudek, I., & Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the national academy of sciences of the United States of America*, *113*(31), 8777–8782. https://doi.org/10.1073/pnas.1601827113

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008a). First aid for depression: a Delphi consensus study with consumers, carers and clinicians. *Journal of affective disorders*, *105*(1-3), 157–165. https://doi.org/10.1016/j.jad.2007.05.004

Langlands, R. L., Jorm, A. F., Kelly, C. M., & Kitchener, B. A. (2008b). First aid recommendations for psychosis: using the Delphi method to gain consensus between mental health consumers, carers, and clinicians. *Schizophrenia bulletin*, *34*(3), 435–443. https://doi.org/10.1093/schbul/sbm099

Linstone, H. A. & Turoff, M. (1975) Delphi Method: Techniques and Applications. *Journal of marketing research, 13*(3), 317-318. https://doi.org/10.2307/3150755

Muris, P., & Ollendick, T. H. (2023). Contemporary Hermits: A Developmental Psychopathology Account of Extreme Social Withdrawal (Hikikomori) in Young People. *Clinical child and family psychology review*, *26*, 459-481. https://doi.org/10.1007/s10567-023-00425-8

Murray, T. J., Pipino, L. L., van Gigch, J. P. (1985). A pilot study of fuzzy set modifications of Delphi. *Humans Systems Management, 5*(1), 76-80. https://doi.org/10.3233/HSM-1985-5111

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World journal of methodology*, *11*(4), 116–129. https://doi.org/10.5662/wjm.v11.i4.116

Niederberger, M., Schifano, J., Deckert, S., Hirt, J., Homberg, A., Köberich, S., Kuhn, R., Rommel, A., Sonnberger, M., & the DEWISS network. (2024). Delphi studies in social and health sciences—Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS ONE, 19*(8), e0304651. https://doi.org/10.1371/journal.pone.0304651

O'Cathain, A., Croot, L., Duncan, E., Rousseau, N., Sworn, K., Turner, K. M., Yardley, L., & Hoddinott, P. (2019). Guidance on how to develop complex interventions to improve health and healthcare. *British medical journal open, 9*(8), e029954. https://doi.org/10.1136/bmjopen-2019-029954

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British medical journal (clinical research ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Popay, J., Roberts, H. M., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Britten, N. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *Institute for health research.* DOI: 10.13140/2.1.1018.4643

Rief, W., Asmundson, G. J. G., Bryant, R. A., Clark, D. M., Ehlers, A., Holmes, E. A., McNally, R. J., Neufeld, C. B., Wilhelm, S., Jaroszewski, A. C., Berg, M., Haberkamp, A., Hofmann, S. G., & PsyChange consortium (2024). The future of psychological treatments: The Marburg Declaration. *Clinical psychology review*, *110*, 102417. https://doi.org/10.1016/j.cpr.2024.102417

Santaguida, P., Dolovich, L., Oliver, D., Lamarche, L., Gilsing, A., Griffith, L.E., Richardson, J., Mangin, D., Kastner, M. & Raina, P. (2018). Protocol for a Delphi consensus exercise to identify a core set of criteria for selecting health related outcome measures (HROM) to be used in primary health care. *Biomed central (family practice), 19,* 152. https://doi.org/10.1186/s12875-018-0831-5

Scheibe, M., Skutsch, M., Schofer, J. (1975). Experiments in Delphi methodology. *In: Linstone H, Turoff M, editors. The Delphi Method: Techniques and Applications,* 262–287. Access: https://www.researchgate.net/publication/242678179_IVC_Experiments_in_Delphi_Methodology

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS medicine*, *8*(1), e1000393. https://doi.org/10.1371/journal.pmed.1000393

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., Moore, L. (2021). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *British medical journal, 374.* https://doi.org/10.1136/bmj.n2061

Sprenkle, D. H., & Piercy, F. P. (Eds.). (2005). *Research methods in family therapy* (2nd ed.). The Guilford Press. Access: https://www.guilford.com/books/Research-Methods-in-Family-Therapy/Sprenkle-Piercy/9781572309609?srsltid=AfmBOopJadp0VD-XNjrdL2B1HsraF9mEUcWJXz8P0fXTGYwncGBJtuT_

Szuhany, K. L., & Simon, N. M. (2022). Anxiety Disorders: A Review. *JAMA*, *328*(24), 2431–2445. https://doi.org/10.1001/jama.2022.22744 von der Gracht, H. A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological forecasting and social change, 79*(8), 1525-1536. https://doi.org/10.1016/j.techfore.2012.04.013

Chapter 2

An exploration of expert consensus views on intervention targets, treatment elements, and key outcomes for extreme social withdrawal (Hikikomori): A Delphi Study

Prepared in accordance with the author requirements for Transcultural Psychiatry

Online Link to Submission Guidelines

Plain Language Summary

Introduction:

'Hikikomori' is a Japanese word that describes people who socially isolate for over six months and rarely interact with others around them. There are many reasons invoked to explain this presentation: mental health difficulties such as depression; feeling a sense of failure; stress from society. Despite the growing recognition of Hikikomori, there is little help available.

Research Questions:

This study asked people with Hikikomori, their parents/carers, and professionals:

1. What problems do you want a treatment to help with?
2. What could a treatment include?
3. How would a treatment be delivered?
4. What goals would a treatment help work towards?

Methods

Participants were grouped based on their experience into those with personal experience of Hikikomori; their parents/carers; and professionals. All participants were surveyed repeatedly between April and November 2024. Their responses were refined over repeat cycles of enquiry as a way of developing a consensus view. Based on their responses, some survey items were removed if statistical analysis indicated that all three groups agreed they were *very important* or *definitely not important*. If participants were not sure, or ratings were in in the middle of the scale, they were re-rated in the next survey. To help with re-rating, all participants were given feedback about how the three groups had rated the item in the most recent round. Some new survey items were also created by the participants after survey one as there was space for them to add comments or share ideas that had been missed by the research team.

Results

Out of 139 survey items, four were agreed to be important by all groups: a treatment should focus on reducing social isolation; a treatment should target socially prescribed perfectionism; peer support should be offered for family members and carers; and that treatment should continue until the individual with Hikikomori can maintain age-appropriate roles independently. Two items were agreed to be not important by all groups: that a treatment should stop after a certain number of sessions/contacts; and that a treatment should stop after a certain number of hours. Individuals with personal Hikikomori experience and their parents/carers were more likely to agree with each other, rather than with professionals.

Conclusion

The results showed large differences in the priorities of the three groups. Future research might want to look into this difference with a larger group of participants. This study is a first step in designing a treatment that listens to the preferences of experts in Hikikomori. Research could potentially build on the identified options and may wish to include these priorities in treatment development.

Abstract

Introduction: Hikikomori is a form of extreme social withdrawal that is now recognised globally. Despite the increased prevalence and substantial impact on the "patient" and their families, evidence-based treatments are of limited quality and efficacy.

Method: To guide treatment development, a Delphi method was used to systematically obtain the opinions of three key stakeholder groups: those with lived experience of Hikikomori (n=2); their parents and carers (n=14); and qualified professionals (n=6). Across four iterative survey rounds, individual and group ratings were analysed against pre-defined thresholds for consensus.

Results: Four items met consensus criteria for inclusion in treatment development: targeting social isolation and socially prescribed perfectionism; the offering of peer support for family and carers; and the importance of Hikikomori maintaining age-appropriate roles. Two items were excluded due to agreed non-importance: the cessation of sessions following a set number of hours or contacts.

Discussion: The results showed substantial variation between the perspectives of the three panels, and significant recruitment difficulties; demonstrating the challenges of involving stakeholders in intervention development. Specifically, the professional panel had conflicting priorities in comparison to the other two groups. Future research may wish to analyse this discrepancy within a larger group, and intervention developers might consider the treatment targets, elements and measurable outcomes favoured by the panels.

Funding: *No funding was required for this review.*
Registration: *This study proposal was registered on Open Science Framework in June 2024: https://osf.io/cjr3u/?view_only=aeb299aaf1954eee92c2742d48e8ddda.*
Statements and Declarations*: The author(s) have no completing interests to declare.*
Key words: *Delphi Method; Hikikomori; Intervention Development; Treatment Development; Stakeholders*

Introduction

Hikikomori

Hikikomori is a Japanese term used to describe significant social and physical isolation and applies both to the condition and the individual living with the experience (Kato et al., 2020). Hikikomori initially emerged in Japan, with the term coming into common use in the late 1990s (Li & Wong, 2015). The lifetime prevalence within Japan was estimated to be 1.2% (Koyama et al., 2010), however Hikikomori is now considered a global phenomenon with reports across Asia, Europe and America (Kato et al., 2018; Krieg & Dickie, 2013). As prevalence increases, the definition has been refined but there are still debates about critical thresholds for the duration of withdrawal and level of social connectedness (Teo & Gaw, 2010; Hattori, 2006). Generally, to meet criteria, an individual must now be completely secluded from society, with significant impacts on daily functioning, for a minimum of six months (Kato et al., 2020). The functional impact of Hikikomori places a social and economic burden on families, carers and society (Li & Wong, 2015). The severity of such impact is often defined as mild to severe, based on the frequency of Hikikomori being able to leave their home or bedroom (Kato et al., 2019).

To complicate matters, the behavioural phenotype of Hikikomori is suspected to mask several possible subtypes that likely have different medical, sociological and medical-anthropological mechanistic drivers (Muris & Ollendick, 2023). Currently a distinction made by clinicians is between primary and secondary Hikikomori. *Primary* Hikikomori is characterised by extreme social withdrawal in the absence of other psychiatric disorders, and rather draws on other various aetiological hypotheses to explain the development of the presentation (Teo & Gaw, 2010). From an attachment theory perspective, withdrawal can be considered as an avoidant coping strategy deployed by individuals with insecure attachment (Tateno et al., 2012). Other developmental perspectives are relevant to disrupted or delayed psychosocial development (Teo et al., 2013). Others place emphasis on factors such as over-dependence due to protective parenting, maladaptive interdependence or counter-dependence (Teo, 2010; Li & Wong, 2015). In contrast, *secondary* Hikikomori describes the withdrawal as a consequence of diagnosable psychiatric disorders such as schizophrenia, social anxiety disorder and avoidant personality disorder (Teo & Gaw, 2010; Hattori, 2006; Nagata et al., 2013). While the withdrawn phenotype can be present across many mental health conditions, the underlying mechanistic drivers, alongside other factors (e.g. complexity and duration), is highly variable and, as such, this impacts treatment approaches and outcomes.

Existing interventions

While it is agreed that Hikikomori warrants therapeutic attention, there are shortcomings both in terms of what is offered, and what is accepted (Li & Wong, 2015). The specific presentation subtype, the potential to accept online peer support, and the motivation to remain Hikikomori can all impact engagement with formal help (Koyama et al., 2010; Sakamoto et al., 2005). Furthermore, some literature cites social stigma and lack of psychoeducation as barriers to families seeking support on behalf of their children (Tateno et al., 2012).

Conventional therapeutic methods used to address extreme social withdrawal include pharmacotherapy (Nagata et al., 2013; Teo, 2010) and counselling (Furlong, 2008; Nagata et al., 2013; Wilson, 2010). Several intervention programmes, and means to assess their efficacy, have also been developed. One example is the family intervention programme based on Mental Health First Aid (MHFA) and Community Reinforcement and Family Training (CRAFT; Kubo et al., 2020; Kubo et al., 2023). A recent randomised controlled trial examined the impact of this 3-day family programme (Kubo et al., 2023). While perceived skills of the family members increased temporarily, the intervention appeared to have no, or negative, effects on other outcome measures including carer confidence, Hikikomori activity levels, attitude and aggressive behaviours. While the study was discontinued due to the pandemic, the preliminary findings highlight the requirement for further development work (Kubo et al., 2023). A second method of supporting Hikikomori involves strengths-based approaches (Matsuguma & Niemiec, 2021). Both cases presented in this study reported a reduction of psychological distress, and an increased use of their strengths, increased self-esteem and increased self-efficacy, assessed across a range of measures (Mewton et al., 2016; Govindji & Linley, 2007; Rosenberg, 1965; Schwarzer & Jerusalem, 1995). Despite promising findings, further work is required to clarify and empirically investigate the approach.

The interventions available vary in terms of delivery mode (e.g. online, face-to-face, hybrid). There are uncertainties around the role of technology in treating Hikikomori, with some warnings of the potential for reliance on technology to worsen isolation (Tateno et al., 2019; Yap et al., 2021), and others highlighting the increased likelihood of some form of social interaction. Used the right way, it could be argued that technology provides a sense of safety and connection and allows Hikikomori to build their self-esteem (Stenseng et al., 2021). For example, the mobile app ‘Pokémon Go' is hoped to encourage those with Hikikomori to go outside without external pressures; emphasising the importance of self-motivation (Kato et al., 2017). Most recently, in a netnography[3] study involving naturalistic observations of an online Hikikomori forum, findings suggested that technology engagement could encourage gradual reintegration. Specifically, factors such as the ability to anonymously tell stories, the wide information access for skill growth and peer networking could support Hikikomori to re-enter society (Park & Yap, 2024).

Development of interventions

The interventions used to support recovery from Hikikomori have mostly drawn upon and adapted pre-existing programmes (e.g. MHFA, CRAFT, Kubo et al., 2023). The purpose of this study was to take a more “bottom up” approach to intervention development, in line with recent shifts away from traditional protocol-for syndrome approaches (Hayes et al., 2020). In essence, a more recently conceptualised process-based approach focuses more on the specific emotional, cognitive and biological mechanisms that drive aspects of psychological disorders. Identifying the relevant psychological mechanisms and processes in the early phases of intervention development may allow for more personalised and targeted treatments, with improved outcomes (Ehring et al., 2022; Rief et al., 2024; Holmes et al., 2018). For example, not only considering the processes underpinning an individual’s presentation, but also the mediators that might impact treatment response (Rief et al., 2024). This is particularly relevant in the field of Hikikomori, where there are often differing comorbidities, severities and underpinning aetiology between individuals.

To capture the variety of perspectives within those affected by Hikikomori, a multi-stakeholder involvement approach was adopted. The updated Medical Research Council (MRC) complex interventions framework encourages stakeholder engagement as a core element of each phase of intervention development in the form of consultation or co-production (Skivington et al., 2021). Core guidance advocates for collaboration and shared decision-making on how to make an intervention accessible, and as such, increase efficacy.

---

[3] Netnography: A qualitative research method used to study online communities through naturalistic observations, where interactions and behaviours can then be analysed (Kozinets & Gambetti, 2020).

Intervention development is advised to be iterative, creative and dynamic, with continual feedback from relevant stakeholders (O'Cathain et al., 2019). While all feedback is informative, structured engagement of stakeholders often yields more reliable output, and as such, consensus-based studies are a beneficial approach (Sinha et al., 2011; Kurvers et al., 2016).

The Delphi method aligns with guidance for complex intervention development as it systematically assembles stakeholders' views through iterative rounds of surveys (Jorm, 2015; Skivington et al., 2021). Furthermore, the Delphi method is particularly helpful in areas of limited-evidence base, diverse presenting problems, or differing opinions (King et al., 2021). This study drew on an international sample including those with Hikikomori experience themselves, their parents/carers and professionals. Analysing and developing consensus on treatment features, between both experts by profession and experts by experience, can allow for a deeper understanding of Hikikomori and enhance the quality and efficacy of treatment development (O'Cathain et al., 2019; Sinha et al., 2011). Furthermore, providing space to share such preferences may enhance the uptake of a care package for a population often described as "hard to engage" (Su et al., 2021).

The current study

This study aimed to systematically capture the key treatment needs and preferences of people impacted by Hikikomori personally and professionally. Consensus was sought on the core processes involved in the development and maintenance of Hikikomori; alongside particular components that are viewed as important for a treatment. Furthermore, identification of meaningful outcomes will streamline practice and enhance engagement. Key stakeholders were asked to consider:

1) *What core psychological processes contribute to Hikikomori generally and which should be targeted in a psychological intervention?*
2) *What accessible and efficient intervention elements would address the identified needs?*
3) *What intervention delivery methods and formats are preferred?*
4) *What intervention outcomes would be relevant, realistic and meaningful?*

Methodology

Registration and Ethics

Ethics approval for this study was obtained from the University of Glasgow College of Medical, Veterinary and Life Sciences (MVLS) Ethics Committee in February 2024 (reference number: 200230210; see Appendix 2.3). This study proposal was registered on Open Science Framework in June 2024:
https://osf.io/cjr3u/?view_only=aeb299aaf1954eee92c2742d48e8ddda (see Appendix 2.2).

Study Design

A four-round, online Delphi method (Jorm, 2015) was conducted to systematically assemble expert opinions regarding Hikikomori treatment. Primarily ordinal data was gathered from each successive survey and analysed against pre-defined thresholds for consensus (see Appendix 2.5). The study was directed by the lead researcher and academic supervisor, with consultation from two experts in the field. A reporting quality checklist guided methodological choices and the write-up (see Appendix 2.1).

Participants

Subject matter experts (SMEs) formed three panels to mirror the diversity of those supporting, and those impacted by, Hikikomori. The Hikikomori experienced panel (HK) required the self-identification of current or historic Hikikomori within the last two years. The parent or carer panel (PC) included families or carers of Hikikomori, or an individual with a history of Hikikomori in the last two years. The professionals panel (PR) involved individuals with relevant qualifications (e.g. training or higher education) and either experience of working with two or more Hikikomori and/or their family systems, or involvement in two relevant publications.

Eligibility criteria included being at least 18 years of age, with internet access, and sufficient command of English to consent and meaningfully participate. Participants were excluded if they self-identified any risk to themselves or others in completing the surveys (see Appendix 2.9).

Recruitment

Recruitment was directed by the lead researcher and academic supervisor. Potential participants were identified through purposive sampling via literature searching and use of existing networks. The study advertisement was shared on social media pages, and directly sent to seventy-nine professionals. Additionally, individuals were invited to share the advertisement with relevant members of their networks. There is no standard requirement for sample size, however 10-18 participants are suggested (Okoli & Pawlowski, 2004). In anticipation of a 20% decline to participate rate, and a 75% response rate for each survey round, we pre-specified a recruitment target of 159 total panel members (Santaguida et al., 2018).

Procedure and Materials

Individuals who responded to the advertisement and completed the confirmation of interest and suitability form were contacted via email with easy read versions of the study outline, demographic and consent information forms (see Appendix 2.4). These documents outlined the study background and objectives, and the research procedure. Information about consent was shared, with an understanding that commencement of the first survey confirms consent to participate. Then, participants were asked to share demographic information and further details of their expertise so that they could be screened for eligibility.

For each successive round of the Delphi surveys, participants were contacted by email with their unique ID code, to remain anonymous to one another, and a hyperlink to the relevant online survey (Qualtrics.com). SMEs were given a deadline of six weeks for each survey, with email reminders every fortnight. Participants had three opportunities to ‘opt in’ to a random prize draw for a £50 Amazon voucher.

The first survey was developed following a literature review, alongside consultation with experts on Hikikomori. Those involved in survey development did not have voting rights in the Delphi panels. Due to resource limitations, a systematic review and a pilot were not feasible, and the surveys were in the English language only. Survey content corresponded to the study aims, whereby participants were invited to rate items under the sub-headings: treatment targets, treatment elements, treatment delivery, and treatment outcomes. In total, Survey 1 contained 136 items and all items across surveys involved 7-point Likert Scales to reflect relative importance (see Appendix 2.4). Survey 1 had open-field text boxes to allow for comments and new suggestions from SMEs, that were analysed and included in Survey 2.

Following completion of each round, Likert ratings were analysed descriptively based on pre-determined criteria for ‘agreement’ and ‘consensus’ (see Results; and Appendix 2.5). Based on this analysis, items were either included in the final consensus statement, removed, or re-rated in the following survey round. Within subsequent rounds, participants received a personalised survey, where each item was presented alongside the participant’s prior rating, and the average rating from each panel. They were then given an opportunity to maintain or revise their rating.

Between April 2024 and November 2024, four Delphi rounds were conducted. A hierarchical stopping procedure determined the end of data collection (Beiderbeck et al., 2021; von Briel, 2018). This was based on stability statistics and a pre-determined rule that no more than four surveys would be completed in the time available.

Results

Subject Matter Expert Demographics

Despite extensive effort, an extension of the recruitment period (10 weeks), and continual monitoring of response rates, 32 individuals registered their interest to participate (20PC, 5HK, 7PR). Of these, 27 registered their consent (an 84.4% response rate; 16PC, 3HK, 8PR), and from this 22 then completed the first survey (an 81.5% response rate; 14PC, 2HK, 6PR; 1 withdraw, 2 did not start, 2 did not finish). All demographic information for Survey 1 participants is presented in Table 2.1.

Table 2.1: Demographic information for participants completing survey 1

	Total N	**Age**	**Country**	**Gender**	**English confirmed as a first language (n, %)**
Parents/ Carers	14	45-52, n=5 55-64, n=7 65-74, n=2	UK, n=13 Ireland, n=1	Female, n=12 Male, n=2	14, 100%
Hikikomori	2	18-24, n=1 25-34, n=1	USA, n=2	Female, n=1 Male, n=1	2, 100%
Professionals	6	25-34, n=2 35-44, n=2 45-54, n=1 65-74, n=1	UK, n=2 USA, n=1 Canada, n=1 Japan, n=1 Italy, n=1	Female, n=2 Male, n=4	2, 33.33%
Total	**22**	**18-24, n=1** **25-34, n=3** **35-44, n=2** **45-54, n=6** **55-64, n=7** **65-74, n=3**	**UK, n-15** **Ireland, n=1** **USA, n=3** **Canada, n=1** **Japan, n=1** **Italy, n=1**	**Female, n=15** **Male, n=7**	**18, 81.81%**

Please see Table 2.2 for retention rates across each round of surveys. Throughout the iterative process, only three individuals (1PC, 2PR) contacted to withdraw their participation due to language difficulties, having limited time, or forgetting to complete the survey.

Table 2.2: Response rate for each panel across the study

	Completed Survey 1	**Completed Survey 2**	**Completed Survey 3**	**Completed Survey 4**
Parents/Carers	14	12	9	9
Professionals	6	5	3	2
Hikikomori	2	2	2	2
Total	**22**	**19**	**14**	**13**
Retention rate	*N/A*	*86.36%*	*73.86%*	*92.86%*

Round 1: Qualitative Analysis

Survey 1 contained free-text boxes to allow for qualitative feedback that was then analysed and used to inform Survey 2 content. A syntax evaluation was conducted to evaluate response types (see Table 2.3). An independent reviewer analysed the first five responses blindly, exceeding the pre-defined threshold of 80% for agreement (92.8%; Lombard, Snyder-Duch & Bracken, 2002).

Content analysis was also conducted (see Table 2.4). Qualitative responses were coded against pre-defined categories derived from Förster and von der Gracht (2014). An independent reviewer analysed the responses of five participants, with a percentage agreement of 71.4%. As such, such, conflicts were discussed and criteria refined. Following this, five additional participant responses were then reviewed, meeting 82.5% agreement between raters, and exceeding the recommended 10% of total responses to be reviewed independently (Lombard et al., 2002).

Table 2.3: Syntax analysis following survey 1

	Hikikomori		**Parents/ Carers**		**Professionals**		**Total**	
Syntax	**Frequency**	**%**	**Frequency**	**%**	**Frequency**	**%**	**Frequency**	**%**
Whole sentences	4	100.0%	44	83.0%	14	73.7%	62	81.6%
Phrases	0	0.0%	9	17.0%	5	26.3%	14	18.4%
Catchwords	0	0.0%	0	0.0%	0	0.0%	0	0.0%

Table 2.4: Content analysis following Survey 1. Hikikomori panel (HK); parent/carer panel (PC); professional panel (PR)

	HK		**PC**		**PR**		**Total**	
Type of Argument	**Frequency**	**%**	**Frequency**	**%**	**Frequency**	**%**	**Frequency**	**%**
Specific instances/cases	0	0.0	0	0.0	0	0.0	0	0.0
Figures	0	0.0	0	0.0	0	0.0	0	0.0
Trends	0	0.0	0	0.0	0	0.0	0	0.0
(Historical) analogies	0	0.0	0	0.0	0	0.0	0	0.0
Developments	0	0.0	0	0.0	0	0.0	0	0.0
Beliefs	3	75.0	21	38.9	8	42.1	32	42.7
Experiences	0	0.0	18	33.3	1	5.3	19	25.3
Differentiation	0	0.0	5	9.3	7	36.8	12	16.0
Lack of information	1	25.0	3	5.6	0	0.0	4	5.3
Misunderstandings	0	0.0	1	1.9	1	5.3	2	2.7
Feedback	0	0.0	1	1.9	2	10.55	3	4.0
Other	0	0.0	5	9.3	0	0.0	5	6.7

Based on the qualitative information gathered in Survey 1, one item was removed due to participants reporting the duplication of an items, and four new items were added:

- *An intervention should consider the role of diet and exercise*
- *An intervention should consider other health issues that may cause withdrawal (e.g. skin disorders or body image difficulties)*
- *Parents should be encouraged to reduce demand (e.g. limit exposure-based tasks) and play a more distanced supportive role*
- *The use of an online portal containing education and advice could be beneficial (e.g. coping strategies)*

Round 1: Quantitative Analysis

There are no universally agreed methods for determining consensus in Delphi methodology. For this study we adapted methods from previous published work to assess agreement and consensus using pre-defined within-panel thresholds:

- *≥75% rating 1-2 (not important items) or 6- 7 (important items) on the Likert scale*
- *A median score of ≤2 (not important items) or ≥6 (important items; Courtenay et al., 2018)*
- *An inter-quartile range (IQR) of ≤1.75 (Beiderbeck et al., 2021)*

An item must have met agreement criteria for importance or unimportance **across all three respondent panels**, to meet complete consensus. Complete consensus resulted in immediate removal of the item from the surveys. If an item met agreement between **two panels** on its **final round of rating** it was considered to have reached partial consensus. These items were appropriate for re-rating, or removal if stability thresholds were met (Koo & Li, 2016). If an item met neither agreement criteria, it was automatically included in the next survey for re-rating. All survey rounds followed this analysis and processing method. A detailed illustration of this process is presented in Figure 2.1.

From the 136 items in Survey 1, no items met complete or partial consensus criteria.

Round 2

In round 2, the adapted survey included 139 items. This round, four items met complete consensus criteria for being "very important":

- *A treatment should target the reduction of social isolation*
- *A treatment should target socially prescribed perfectionism (defined as: perceived social pressures and expectations to be perfect in the eyes of others)*
- *The offering of peer support for family members and carers of Hikikomori is important*
- *A meaningful outcome to indicate that treatment is being effective, is when the individual with Hikikomori maintains age-appropriate roles independently (e.g. work or study)*

Additionally, two items met complete consensus criteria for being unimportant:

- *A treatment should stop when a certain number of sessions/contacts have been had*
- *A treatment should stop when a certain number of hours have been used*

Eleven items met criteria for partial consensus for importance between the PC and HK panels only:

- *A treatment should target the experience of defeat or viewing oneself as a failure*
- *A treatment should focus on reducing feelings of hopelessness*
- *A treatment should target poor self-esteem, or a lack of self-confidence*
- *A treatment should target low motivation*
- *A treatment should target a lack of belonging or purpose*
- *A treatment should focus on self-stigma (defined as: internalised shame about difficulties or differences*
- *A treatment should involve the development of distress tolerance skills (i.e. helping an individual manage distress and discomfort)*

- *A treatment should include the development of self-understanding and reflective abilities (e.g. through metacognitive work (defined as: awareness and understanding of one's own thought process and the mind of others)*
- *The inclusion of cognitive restructuring (thought identification, challenging beliefs and modifying negative cognitions) is important*
- *Treatment should include the training of mental health first aid for families/carers (e.g. active listening)*
- *A sign a treatment has been effective is when the individual with Hikikomori maintains age-appropriate roles (e.g. work or study) with support*

Two items met criteria for partial consensus of importance between the PC and PR panels:

- *There should be the provision of general family support alongside professional intervention*
- *A treatment should include receiving psychoeducation about Hikikomori*

In line with the pre-defined hierarchical stopping criteria, stability of ratings across rounds was assessed using the intra-class correlation coefficient (ICC; two-way mixed model, single-rater, absolute agreement). Coefficients of good or excellent (i.e. <.75) were accepted as significant, and such items were removed from subsequent surveys (Koo & Li, 2016). In Survey 2, stability statistics led to the removal of 68 items (see Appendix 2.8).

Round 3

Survey 3 included 65 items. Within this round, zero items met complete consensus criteria. Three items met criteria for partial consensus between the PC and HK panels only:

- *The offering of peer support for Hikikomori (e.g. buddies, mentors) is important*
- *There should be an offering of de-stigmatising workshops and events*
- *A sign that the treatment is being effective is when the individual with Hikikomori demonstrates willingness to participate socially*

Stability statistics led to the removal of 40 items (see Appendix 2.8).

Round 4

The final survey consisted of 25 items. No items met complete consensus criteria. Two items met partial consensus criteria for importance, and were agreed between the PC and HK panels only:

- *A treatment should include problem-solving skills training*
- *Observations by family could be used to assess treatment change*

Additionally, stability statistics led to the removal of 15 items (see Appendix 2.8). Based on the hierarchical stopping procedure, data collection ended after this survey. At the final round, there were 10 items that did not meet stability or agreement criteria (see Appendix 2.8).

Summary

In summary, across all four surveys, four items met complete consensus criteria for importance:

- *A treatment should target the reduction of social isolation*
- *A treatment should target socially prescribed perfectionism (defined as: perceived social pressures and expectations to be perfect in the eyes of others)*
- *The offering of peer support for family members and carers of Hikikomori is important*
- *A meaningful outcome to indicate that treatment is being effective, is when the individual with Hikikomori maintains age-appropriate roles independently (e.g. work or study)*

Across the surveys, two items met complete consensus criteria for unimportance:

- *A treatment should stop when a certain number of sessions/contacts have been had*
- *A treatment should stop when a certain number of hours have been used*

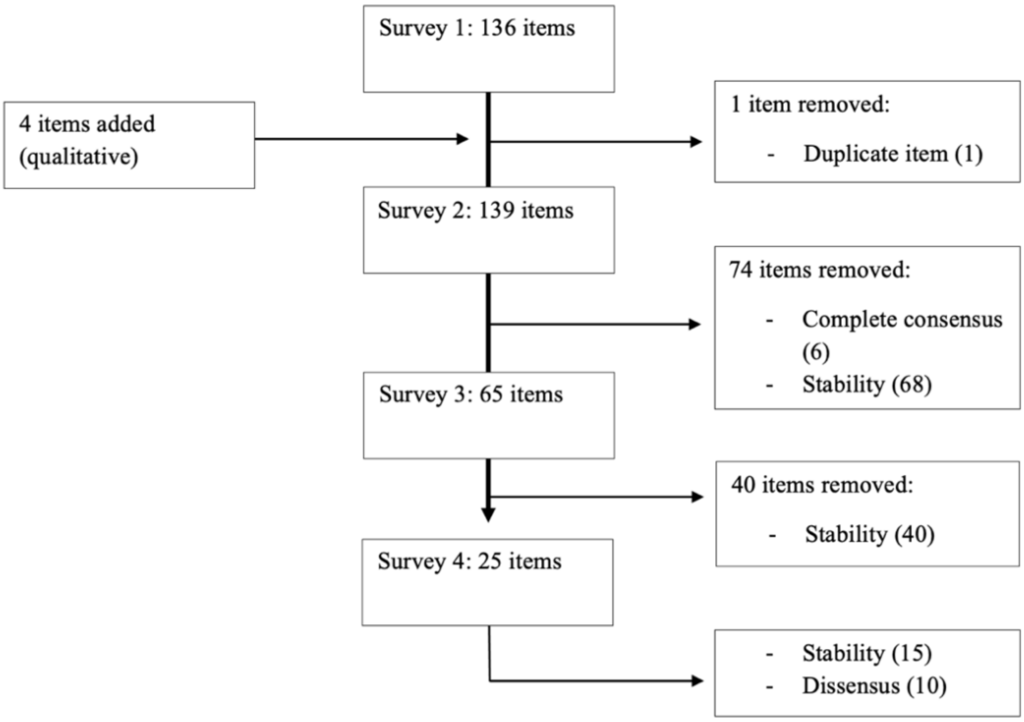

Figure 2.1: Overview of the survey items, consensus and stability

Discussion

The aim of the current study was to use consensus-building methods to ascertain key aspects of a treatment for Hikikomori. Relevant stakeholders were asked to rate and contribute new items relevant to the subcategories of treatment targets, treatment elements, delivery methods and meaningful outcomes (as discussed below). Despite six items meeting complete consensus, there was substantial variation in the perspectives of the three panels; illustrating the challenge of meeting agreement for Hikikomori. Furthermore, the pattern of engagement suggests that motivation to participate may be influenced by the perceived impact of the withdrawal (for instance, families and carers having a stronger vested interest in the development of a treatment).

Treatment Targets

Respondents were asked to rate the importance of 40 treatment targets (two of which were new stakeholder additions) across four rounds. All three panels agreed that a treatment for Hikikomori should target the reduction of social isolation (as was expected due to the nature of Hikikomori) and socially prescribed perfectionism. Socially prescribed perfectionism being highly endorsed as a key issue lends itself well to some findings that the risk of withdrawal is increased due to parental pressures and expectations (Umeda & Kawakami, 2012; Wong et al., 2019). The parent/carer and Hikikomori panels agreed that treatment should target: the experience of defeat or viewing oneself as a failure; reducing feelings of hopelessness; poor self-esteem or a lack of self-confidence; low motivation; a lack of belonging or person; and self-stigma. Many of these aspects could be co-dependent with one another, with the 'view of the self' having a significant feature. When examining Japanese youths who were of high risk of developing Hikikomori, an "entity view of self" rather than an "incremental view of self" played a role in their responses to failure (Norasakkunkit & Uchida, 2014). Existing literature demonstrates awareness of these factors, as self-esteem and self-confidence are frequently used within measurement scales, or as interventions targets; therefore, it is unexpected that the professional panel did not recognise the importance of this (e.g. Uchida & Norasakkunkit, 2015; Kubo et al., 2023). With regards to reduced self-motivation, some literature suggests that less directive interventions may be more accessible for Hikikomori re-entering society (e.g. 'Pokémon Go'; Kato et al., 2017). Less directive interventions that mirror ‘play’ (a more accessible way of exploring the world) may reduce the likelihood of such internal and external pressures arising (e.g. criticism from others, self-judgement).

Treatment Elements

Participants were then asked to rate the relative importance of 32 treatment elements, and only one item met criteria for importance across all three panels: the inclusion of peer support for family members and carers of Hikikomori. Hikikomori can have a widespread impact on family members, impacting mood, marital and psychological conflict (Funakoshi & Miyamoto, 2014). The endorsement of this item from all panels indicates the awareness of such strain within families. The provision of support for families would not only minimise this impact but may have a knock-on effect for the individual with Hikikomori. Dysfunctional dynamics, parental attitudes and stigma impact withdrawn behaviours, and these may all be heightened due to caregiver burnout (Li & Wong, 2015). As such, it could be hypothesised that support for family members may increase their wellbeing and decrease dynamics that may be contributing to, or maintaining, the social withdrawal.

Seven items were agreed between the parent/carer and Hikikomori panels only, again, highlighting the divergent views displayed by the professional panel. These items included: the development of distress tolerance skills; the development of self-understanding and reflective abilities; problem-solving skills training; cognitive restructuring; peer support for Hikikomori; MHFA training for families/carers; the offering of de-stigmatising workshops and events. Some of these aspects are in line with Kubo and colleague’s (2023) work that includes MHFA training for parents, with measures of outcome involving stigma and perceived skills. The lack of endorsement on this technique by professionals may indicate a shift away from this method or could demonstrate the disparity in professional judgements. The professional panel also did not endorse peer support for Hikikomori as an item of high importance. While Hikikomori can find peer support more beneficial than professional guidance, there may be concerns that inaccurate information could be shared that could be harmful for Hikikomori (Haasio et al., 2023).

Two items were agreed between the parent/carer and professional panels only. These items involved: the provision of general family support alongside professional intervention; and involving psychoeducation about Hikikomori. The divergence in agreement may indicate that Hikikomori view peer support for their family members as more valuable than general family support, due to their own positive experiences of this (Haasio et al., 2023). Additionally, online forums contain extensive information about Hikikomori, and as such, they may feel that further education is not a priority or necessary (Park & Yap, 2024).

Treatment Delivery

The surveys included 20 items regarding how a treatment would be delivered (one of which was a suggestion following qualitative feedback in Survey 1) and none of these items met criteria for partial or complete consensus. While the parent/carer and professional panels swayed between endorsing face to face treatment across rounds, the Hikikomori panel *never* supported this item. However, they did rate the use of internet forums and blogs at one point during the study. These findings relate to existing literature, where the evidence debates between technology being a tool to reintegrate Hikikomori into society, or something that exacerbates their isolation (Tateno et al., 2019; Yap et al., 2021; Kato et al., 2017). The use of forums in this population is significant (e.g. the 'Hikikomori' group on Reddit has 18,000 members). Hikikomori may experience a sense of community, a space of no-judgement (as often these are anonymous) and validation from shared-understanding with peers (Tateno et al., 2012; Park & Yap, 2024).

Guidance for intervention development emphasises person-centered care, and a need to consider means to maximise engagement (Skivington et al., 2021). However, the systemic impact of Hikikomori also means considering the needs of their parents and carers, and how they may best be supported. The disparity of preferences between all panels warrants further analysis, and it may be that different modes of support are offered to Hikikomori and their carers.

Treatment Outcomes

The participants were also asked to rank 48 items that specified potentially meaningful outcomes in Hikikomori treatment. Within this subcategory, only one item met complete consensus criteria for importance: that treatment efficacy would be demonstrated if the individual with Hikikomori became able to independently maintain age-appropriate roles. Two items also met complete consensus criteria but for unimportance, that a treatment should stop when: a certain number of sessions/contacts have been had; or when a certain number of hours have been used. The response to three items demonstrates the consensus that flexible and patient approaches are required for this population (e.g. to form a therapeutic alliance; Krieg & Dickie, 2013), as opposed to time restricted practices. It is likely that this preference will need to be balanced with the economic climate and current state of health care (i.e. long waitlists and high cost of care). Hikikomori may wish for such flexibility due to fear or uncertainty of consistent engagement, or feeling more comfortable with an open, exploratory approach. It is understood through contact with members of the parent/carer panel, that they feel neglected and unsupported by current care options, and as such, the idea of 'limited' care may have even more threat attached for this panel.

Additionally, three items met importance criteria between the parent/carer and Hikikomori panels only. These items suggested that a sign that a treatment has been effective is when an individual with Hikikomori maintains age-appropriate roles with support, or when they demonstrate willingness to participate socially. These two panels also agreed that observations by family may be useful to assess treatment change. This warrants further consideration by researchers, as perhaps if questionnaires are required, there may be more preference for parental report scales (e.g. Hikikomori Behaviour Checklist; Sakai et al., 2004) than self-report scales (e.g. Hikikomori Questionnaire; Teo et al., 2018). Given that the self-report scales rely on engagement of the Hikikomori, the preference for family observations by the parent/carer and Hikikomori panels is plausible. It could be that professionals value feedback from Hikikomori themselves to increase reflection and insight, and that is why this item was not endorsed by them.

Recruitment and Panel Participation

Intervention development guidance advocates for stakeholder involvement (Skivington et al., 2021) however, this study illustrated the difficulty in recruiting stakeholders for a study about Hikikomori. This may be due to the nature of the socially withdrawn phenotype; the culturally restricted elements of the study (e.g. survey language); or the smaller group of researchers.

Despite extensive efforts to recruit panellists (purposive sampling, snowball sampling, extending the recruitment period, email reminders), recruitment was significantly lower than hoped. Out of 79 contacted professionals, seven registered their interest, six started the study, and two completed it. Despite low numbers and retention, the professional panel was the most diverse in terms of country of residence. The differing cultural perspectives and command of English may have impacted the ease of engagement with the surveys and influenced the limited consensus found within this panel. The recruitment of Hikikomori was an anticipated difficulty and while an advertisement posted on an online Hikikomori forum was viewed by over 3000 people, five registered their interest in participating. The two individuals who did start the survey were retained through the duration of the study. In contrast to the professional and Hikikomori panels, parents and carers were more readily recruited to the study. A Facebook support page played a significant role in this. Twenty parents/carers registered their interest, with fourteen completing survey 1, and nine completing all four rounds.

The recruitment profile for the study warrants further consideration. The significant imbalance of interest for participating in the study suggests something about the motivation across the panels. The challenge of recruiting professionals may indicate limited engagement with the study, potentially reflecting greater rigidity in professionals' perspectives, lack of time to commit to a consensus-study, or language barriers. The parent and carer interest was significantly higher, with several informal requests for support or guidance, or reflections of frustration over the lack of existing help in the UK.

In line with guidance, Survey 1 contained open-field text boxes (Sinha et al., 2011; Jünger et al., 2017). A syntax evaluation was conducted to establish engagement across groups and to serve as a quality measure of responses (Roßmann et al., 2018). Across panels, responses mostly consisted of whole sentences, suggesting high levels of engagement. Indicated by a lower ratio of whole sentences to phrases, it could be interpreted that the professional panel responded with the least detail. Content analysis was also conducted to assess answering behaviour. While all panellists more commonly reported their beliefs in free-text responses, there were differences in the distribution of panels across the pre-defined categories. The professional panel were more likely than other panels to include differentiations (i.e. ‘if, then’ statements). This may suggest that professionals were more likely to take a systematic, nuanced approach to answering and tended to structure their points with logical caveats or conditions. In contrast, the parent/carer group were more likely than others to rely on their experiences to justify their points, using a narrative, potentially emotion-driven, approach to answering. The Hikikomori panel were particularly more likely to include beliefs in their qualitative feedback. Beliefs tend to be values-based and intuitive, and so it appears answers relied on this, as opposed to structured reasoning or personal experiences.

Only six of 139 items were agreed by all panels; however, the absence of consensus can be highly informative (Scheibe et al., 2002). For example, the higher rate of agreement between the parent/carer and Hikikomori panels could suggest a potential mismatch between what professionals offer in terms of treatment and what is up taken by those with Hikikomori. However, these results should be interpreted with caution, as the professional panel was small and the most culturally diverse, so the final output may have been skewed (Förster & von der Gracht, 2014). Additionally, due to the nature of design (i.e. requiring consensus across three separate panels), the overall ‘agreement’ was lower than it would have been in the absence of panels. While increasing the likelihood of ‘agreement’ could be appealing to researchers to increase the final consensus, with high levels of heterogeneity, it is recommended to include panels or detailed analysis to truly ascertain the diversity of perspectives (Beiderbeck et al., 2021). Given the different perspectives, it may be beneficial for researchers to consider dividing the recipients of the treatment (i.e. Hikikomori and parents/carers) to investigate if greater agreement is achieved.

Strengths

This is the first study of its kind regarding Hikikomori, adding to a small evidence base. Involving stakeholders’ opinions through a Delphi method was an invaluable contribution that is in line with best practice for complex intervention development (Skivington et al., 2021). This study used the ACCORD framework to ensure high quality reporting of the methodology, and reduce researcher and reporting bias (Gattrell et al., 2024). The DELPHISTARR (Delphi SelecTing Approaches for Rapid Reviews; Niederberger et al.,

2024) should be considered by future researchers as a method-specific quality appraisal tool, however, this had not been published at the time of developing and structuring the project.

This study not only examined overall agreement between SMEs, but the separation into panels allowed for more accurate understanding of the diverse perspectives that participants shared (Beiderbeck et al., 2021). The anonymous and online manner of the Delphi may have encouraged the participation from those with lived experience, which given the nature of Hikikomori, is a rare and valuable insight.

In addition to the rating of items, qualitative feedback allowed for the opportunity to share practical-based experience, and as such built on the theory-driven initial item set; another practice that is in line with intervention development guidance (Skivington et al., 2021).

While the sample size was low, the response rate (73.86%-92.86%) was above the recommended 70% (Ali et al., 2018), and in keeping with other Delphi studies informing intervention development (Connell et al., 2020). The small sample size meant that the overall percentage agreement could have been influenced easily by few raters, and as such, the analysis included central tendency measures to assess the variability of ratings (Ali et al., 2018). Additionally, calculating the stability allowed a further, more robust assessment into the answering pattern of panellists (Nasa et al., 2021).

Limitations

As is the case with any study, this research is not without limitations. Due to the nature of the Delphi methodology, the findings are specific to the group of recruited participants. Given the nature of Hikikomori, it could be possible that those who volunteered to participate may be unrepresentative of the wider population.

While the online nature of the project and sampling methods took measures to capture a range of expertise and cultures, the participant group remain broadly UK based (68.18%). Additionally, due to resource constraints, the survey was unable to be translated, and as such a sufficient command of English was required. At the initial survey, 81.82% reported that English was their first language. While the goal of the Delphi method is not to provide generalisable results, and rather capture the opinions of the experts participating, the cultural nuances of the Hikikomori presentation must be considered when conducting a study of this methodology. The drivers underlying Hikikomori, and the perception and response of support systems, varies significantly across cultures. As highlighted, the experience and perception of the 'self', in addition to stigma and cultural expectations must be considered when understanding and treating Hikikomori. Therefore, future studies may wish to expand the scope of participants or consider more culturally-specific Delphi studies.

As highlighted, another limitation is the small sample size of the study. Despite reports that the selection of suitable experts should be prioritised over statistical power (Okoli & Pawlowski, 2004), with such a small sample size, results are easily skewed, and each panellist had significant sway. This is particularly the case in the Hikikomori and professional panels, where sample sizes were smaller. Small sample size and high attrition is a common difficulty within Delphi methodology due to the long period of data collection and longer surveys typically near the start of the process (Hsu & Sandford, 2007). While efforts were made to reduce participant burden and fatigue (subcategories, pre-agreed email reminders), it was a significant commitment to engage in all four surveys.

Implications and Future Research

This study is the first step in a dynamic process that captures stakeholder perspectives in Hikikomori intervention development. While results should be interpreted with caution, items that were completely or partially endorsed by the panellists serve as useful information in terms of a future intervention. Researchers may wish to assess the cross-cultural validity of this study's developed consensus. Further work may involve re-assessing consensus in a larger group or considering translating resources to increase the reach of the consensus exercise. Additionally, due to the diversity of perspectives observed, future research may consider gathering consensus again but for two separate classes of treatment recipients (Hikikomori and parents/carers) to investigate if this increases the likelihood for consensus.

In addition to the consensus analysis, the panellists' answering behaviour has provided useful qualitative information in terms of how aligned professionals may be to those with lived experience, alongside the differing levels of motivation or capacity to engage in stakeholder-based research. While recommended as a key component in best practice for treatment development, this study indicates the challenges of recruiting stakeholders; particularly in research involving social withdrawal. Researchers may wish to reduce the length of Hikikomori duration required from 6 months to 3 months to enhance recruitment, as generally there is little difference to presentation (Wong et al., 2015).

Conclusions

This Delphi study aimed to ascertain key stakeholder's opinions on aspects of a treatment for Hikikomori. Based on the findings of this study, a future psychologically informed intervention for Hikikomori would primarily target the reduction of social isolation and socially prescribed perfectionism. While the literature base and outcome measures to assess treatment give reference to self-esteem more broadly, socially prescribed perfectionism has not been the primary focus of an intervention for Hikikomori thus far. It may be that the intervention will target such processes in a transdiagnostic manner, or person-specific processes could be identified via idiosyncratic formulations; and as such, treatments can be adapted for individual needs and cultures. The importance of including peer support for the family members and carers of Hikikomori was highlighted. Additionally, the intervention should not be limited by a number of sessions or hours and rather should consider changes in an individual with Hikikomori's ability to maintain age-appropriate roles independently. The perspectives gathered indicated a preference for more flexible, open support, rather than the implementation of limited care. Future research should consider how to enhance the recruitment of Hikikomori, and thus the quality of the consensus statement, and may wish to further investigate the dissonance between professionals and those with lived experience.

References

Ali, N., Rigney, G., Weiss, S. K., Brown, C. A., Constantin, E., Godbout, R., Hanlon-Dearman, A., Ipsiroglu, O., Reid, G. J., Shea, S., Smith, I. M., Van der Loos, M., & Corkum, P. V. (2018). Optimizing an eHealth insomnia intervention for children with neurodevelopmental disorders: a Delphi study. *Sleep Health*, *4*(2), 224–234. https://doi.org/10.1016/j.sleh.2017.12.008

Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, *8*, 101401. https://doi.org/10.1016/j.mex.2021.101401

Connell, C. J., Plummer, V., Crawford, K., Endacott, R., Foley, P., Griffiths, D.L., Innes, K., Nayna, S. P., Walker, L. E., & Morphet, J. (2020). Practice priorities for acute care nursing: A Delphi study. *Journal of Clinical Nursing, 29*(0), 2615-2625. https://doi.org/10.1111/jocn.15284

Courtenay, M., Deslandes, R., Harries-Huntley, G., Hodson, K., & Morris, G. (2018). Classic e-Delphi survey to provide national consensus and establish priorities with regards to the factors that promote the implementation and continued development of non-medical prescribing within health services in Wales. *British Medical Journal Open, 8*(9), e024161. https://doi.org/10.1136/bmjopen-2018-024161

Ehring, T., Limburg, K., Kunze, A. E., Wittekind, C. E., Werner, G. G., Wolkenstein, L., Guzey, M., & Cludius, B. (2022). (When and how) does basic research in clinical psychology lead to more effective psychological treatment for mental disorders? *Clinical psychology review*, *95*, 102163. https://doi.org/10.1016/j.cpr.2022.102163

Förster, B., & von der Gracht, H. (2014). Assessing Delphi panel composition for strategic foresight - a comparison of panels based on company-internal and external participants. *Technological Forecasting and Social Change, 84,* 215–229. http://dx.doi.org/10.1016/j.techfore.2013.07.012

Funakoshi, A. & Miyamoto, Y. (2014) Significant Factors in Family Difficulties for Fathers and Mothers Who Use Support Services for Children with Hikikomori. *Psychiatry and Clinical Neuroscience, 69,* 210-9. http://dx.doi.org/10.1111/pcn.12230

Furlong, A. (2008) The Japanese Hikikomori Phenomenon: Acute Social Withdrawal among Young People. *Sociological Review, 56,* 309-325.

https://doi.org/10.1111/j.1467-954X.2008.00790.x

Govindji, R., & Linley, P. A. (2007). Strengths use, self-concordance and well-being: Implications for Strengths Coaching and Coaching Psychologists. *International Coaching Psychology Review 2*(2). https://doi.org/10.53841/bpsicpr.2007.2.2.143

Hassio, A., Kannasto, E., & Salminen-Tuomaala, M. (2023). Sharing mental health peer information and peer support in hikikomori's online discussion groups. *Qualitative and Quantitative Methods in Libraries, 12*(3), 329-343. Available at: https://qqml-journal.net/index.php/qqml/article/view/818

Hattori, Y. (2006). Social Withdrawal in Japanese Youth: A Case Study of Thirty-Five Hikikomori Clients. *Journal of Trauma Practice*, *4*(3–4), 181–201. https://doi.org/10.1300/J189v04n03_01

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020). A process-based approach to psychological diagnosis and treatment:The conceptual and treatment utility of an extended evolutionary meta model. *Clinical psychology review*, *82*, 101908. https://doi.org/10.1016/j.cpr.2020.101908

Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., Roiser, J. P., Bockting, C. L. H., O'Connor, R. C., Shafran, R., Moulds, M. L., & Craske, M. G. (2018). The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *The lancet. Psychiatry*, *5*(3), 237–286. https://doi.org/10.1016/S2215-0366(17)30513-8

Hsu, C. & Sandford, B. A., (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research, and Evaluation 12*(1), 10. https://doi.org/10.7275/pdz9-th90

Jorm A. F. (2015). Using the Delphi expert consensus method in mental health research. *The Australian and New Zealand Journal of Psychiatry, 49*(10), 887–897. https://doi.org/10.1177/0004867415600891

Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative medicine*, *31*(8), 684–706. https://doi.org/10.1177/0269216317690685

Kato, T. A., Kanba, S., & Teo, A. R. (2020). Defining pathological social withdrawal: proposed diagnostic criteria for hikikomori. *World Psychiatry, 19*(1), 116-117. https://doi.org/10.1002/wps.20705

Kato, T. A., Kanba, S. & Teo, A. R. (2019). Hikikomori: Multidimensional understanding, assessment, and future international perspectives. *Psychiatry and Clinical Neurosciences, 73*, 427-440. https://doi.org/10.1111/pcn.12895

Kato, T. A., Kanba, S., & Teo, A. R. (2018). Hikikomori: experience in Japan and international relevance. *World Psychiatry, 17*(1), 105. https://doi.org/10.1002/wps.20497

Kato, T. A., Teo, A. R., Tateno, M., Watabe, M., Kubo, H., & Kanba, S. (2017). Can Pokémon GO rescue shut-ins (hikikomori) from their isolated world? *Psychiatry and Clinical Neurosciences*, *71*(1), 75–76. https://doi.org/10.1111/pcn.12481

King, P. R., Jr., Beehler, G. P., Donnelly, K., Funderburk, J. S., & Wray, L. O. (2021). A practical guide to applying the Delphi technique in mental health treatment adaptation: The example of Enhanced Problem-Solving Training (E-PST). *Professional psychology: research and practice, 52*(4), 376–386. https://doi.org/10.1037/pro0000371

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Koyama, A., Miyake, Y., Kawakami, N., Tsuchiya, M., Tachimori, H., Takeshima, T., & World Mental Health Japan Survey Group, 2002-2006 (2010). Lifetime prevalence, psychiatric comorbidity and demographic correlates of "hikikomori" in a community population in Japan. *Psychiatry Research*, *176*(1), 69–74. https://doi.org/10.1016/j.psychres.2008.10.019

Kozinets, R.V., & Gambetti, R. (2020). Netnography Unlimited: Understanding Technoculture using Qualitative Social Media Research (1st ed.). *Routledge.* https://doi.org/10.4324/9781003001430

Krieg, A. & Dickie, J. R. (2013), "Attachment and hikikomori: a psychosocial developmental model", *International Journal of Social Psychiatry*, 59(1), 61-72. https://doi.org/10.1177/0020764011423182

Kubo, H., Urata, H., Sakai, M., Nonaka, S., Saito, K., Tateno, M., Kobara, K., Hashimoto, N., Fujisawa, D., Suzuki, Y., Otsuka, K., Kamimae, H., Muto, Y., Usami, T., Honda, Y., Kishimoto, J., Kuroki, T., Kanba, S., & Kato, T. A. (2020). Development of 5-day hikikomori intervention program for family members: A single-arm pilot trial. *Heliyon*, *6*(1), e03011. https://doi.org/10.1016/j.heliyon.2019.e03011

Kubo, H., Urata, H., Sakai, M., Nonaka, S., Kishimoto, J., Saito, K., Tateno, M., Kobara, K., Fujisawa, D., Hashimoto, N., Suzuki, Y., Honda, Y., Nakao, T., Otsuka, K., Kanba, S., Kuroki, T., & Kato, T.A. (2023). 3-day intervention program for family members of hikikomori sufferers: A pilot randomized controlled trial. *Frontiers in Psychiatry, 13,* 1029653. https://doi.org/10.3389/fpsyt.2022.1029653

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., Argenziano, G., Zalaudek, I., & Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the national academy of sciences of the United States of America*, *113*(31), 8777–8782. https://doi.org/10.1073/pnas.1601827113

Li, T. M., & Wong, P. W. (2015). Youth social withdrawal behavior (hikikomori): A systematic review of qualitative and quantitative studies. *Australian & New Zealand Journal of Psychiatry*, *49*(7), 595-609. https://doi.org/10.1177/0004867415581179

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

Matsuguma, S., & Niemiec, R. M. (2021). Hikikomori from the Perspective of Overuse, Underuse, and Optimal Use of Character Strengths: Case Reports. *International Journal of Applied Positive Psychology*, *6*, 219–231. https://doi.org/10.1007/s41042-020-00047-3

Mewton, L., Kessler, R. C., Slade, T., Hobbs, M. J., Brownhill, L., Birrell, L., Tonks, Z., Teesson, M., Newton, N., Chapman, C., Allsop, S., Hides, L., McBride, N., & Andrews, G. (2016). The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents. *Psychological assessment*, *28*(10), 1232–1242. https://doi.org/10.1037/pas0000239

Muris, P., & Ollendick, T. H. (2023). Contemporary Hermits: A Developmental Psychopathology Account of Extreme Social Withdrawal (Hikikomori) in Young People. *Clinical Child and Family Psychology Review*, *26*, 459-481. https://doi.org/10.1007/s10567-023-00425-8

Nagata, T., Yamada, H., Teo, A. R., Yoshimura, C., Nakajima, T., & van Vliet, I. (2013). Comorbid social withdrawal (hikikomori) in outpatients with social anxiety disorder: clinical characteristics and treatment response in a case series. *The International Journal of Social Psychiatry*, *59*(1), 73–78.

https://doi.org/10.1177/0020764011423184

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, *11*(4), 116–129. https://doi.org/10.5662/wjm.v11.i4.116

Norasakkunkit, V., & Uchida, Y. (2014). To conform or to maintain self-consistency? Hikikomori risk in Japan and the deviation from seeking harmony. *Journal of Social and Clinical Psychology, 33,* 918–935. https://doi.org/10.1521/jscp.2014.33.10.918

O'Cathain, A., Croot, L., Duncan, E., Rousseau, N., Sworn, K., Turner, K. M., Yardley, L., & Hoddinott, P. (2019). Guidance on how to develop complex interventions to improve health and healthcare. *British Medical Journal Open, 9*(8), e029954. https://doi.org/10.1136/bmjopen-2019-029954

Okoli, C. & Pawlowski, S. (2004). The Delphi Method as a Research Tool: An Example, Design Considerations and Applications. *Information & Management, 42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Park, H. E., & Yap, S. F. (2024). Technology affordances and social withdrawal: The rise of hikikomori. *Psychology & Marketing, 41,* 1469–1488. https://doi.org/10.1002/mar.21991

Rief, W., Asmundson, G. J. G., Bryant, R. A., Clark, D. M., Ehlers, A., Holmes, E. A., McNally, R. J., Neufeld, C. B., Wilhelm, S., Jaroszewski, A. C., Berg, M., Haberkamp, A., Hofmann, S. G., & PsyChange consortium (2024). The future of psychological treatments: The Marburg Declaration. *Clinical psychology review*, *110*, 102417. https://doi.org/10.1016/j.cpr.2024.102417

Roßmann, B., Canzaniello, A., von der Gracht, H., Hartmann, E. (2018). The future and social impact of big data analytics in supply chain management: results from a Delphi study. *Technological Forecasting and Social Change; 130*: 135–149. https://doi.org/10.1016/j.techfore.2017.10.005

Rosenberg, M. (1965). Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400876136

Sakai, M., Ishikawa, S.-I., Sato, H. and Sakano, Y. (2004) Development of Hikikomori Behavior Checklist (HBCL) and Examination of Its Reliability and Validity. *Japanese Journal of Counselling Science, 37,* 210-220. https://doi.org/10.1037/t68812-000

Sakamoto, N., Martin, R. G., Kumano, H., Kuboki, T., & Al-Adawi, S. (2005). Hikikomori, is it a culture-reactive or culture-bound syndrome? Nidotherapy and a clinical vignette from Oman. *International Journal of Psychiatry in Medicine*, *35*(2), 191–198. https://doi.org/10.2190/7WEQ-216D-TVNH-PQJ1

Santaguida, P., Dolovich, L., Oliver, D., Lamarche, L., Gilsing, A., Griffith, L.E., Richardson, J., Mangin, D., Kastner, M. & Raina, P. (2018). Protocol for a Delphi consensus exercise to identify a core set of criteria for selecting health related outcome measures (HROM) to be used in primary health care. *BMC Family Practice, 19,* 152. https://doi.org/10.1186/s12875-018-0831-5

Scheibe, M., Skutsch, M. & Schofer, J. (2002). Experiments in Delphi Methodology. In H. A. Linstone & M. Turoff (Eds.), *The Delphi method. Techniques and Applications* (pp. 257-281). Addison Wesley Publishing Company. Available at: https://www.foresight.pl/assets/downloads/publications/Turoff_Linstone.pdf

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35-37. https://psycnet.apa.org/doi/10.1037/t00393-000

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*, *8*(1), e1000393. https://doi.org/10.1371/journal.pmed.1000393

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., Moore, L. (2021). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *British Medical Journal, 374.* https://doi.org/10.1136/bmj.n2061

Stenseng, F., Falch-Madsen, J., & Hygen, B. W. (2021). Are there two types of escapism? Exploring a dualistic model of escapism in digital gaming and online streaming. *Psychology of Popular Media, 10*(3), 319–329. https://doi.org/10.1037/ppm0000339

Su, X., Li, T. M. L., & Wong, P. W. C. (2021). Expanding sources of recognition: Towards an experience-driven framework for reengaging hikikomori and semi-hikikomori. *Child and Adolescent Social Work Journal, 40,* 713-725. https://doi.org/10.1007/s10560-021-00795-4

Tateno, M., Park, T. W., Kato, T. A., Umene-Nakano, W., & Saito, T. (2012). Hikikomori as a possible clinical term in psychiatry: a questionnaire survey. *BMC Psychiatry*, *12*, 169. https://doi.org/10.1186/1471-244X-12-169

Tateno, M., Teo, A. R., Ukai, W., Kanazawa, J., Katsuki, R., Kubo, H., & Kato, T. A. (2019). Internet Addiction, Smartphone Addiction, and Hikikomori Trait in Japanese Young Adult: Social Isolation and Social Network. *Frontiers in Psychiatry*, *10*, 455. https://doi.org/10.3389/fpsyt.2019.00455

Teo, A. R., & Gaw, A. C. (2010). Hikikomori, a Japanese culture-bound syndrome of social withdrawal?: A proposal for DSM-5. *The Journal of Nervous and Mental Disease*, *198*(6), 444–449. https://doi.org/10.1097/NMD.0b013e3181e086b1

Teo, A. R., Stufflebam, K. W., & Kato, T. A. (2013). The Intersection of Culture and Solitude: The Hikikomori Phenomenon in Japan. *The Handbook of Solitude: Psychological Perspectives on Social Isolation, Social Withdrawal, and Being Alone,* 445-460. https://doi.org/10.1002/9781118427378.ch25

Teo A. R. (2010). A new form of social withdrawal in Japan: a review of hikikomori. *The International Journal of Social Psychiatry*, *56*(2), 178–185. https://doi.org/10.1177/0020764008100629

Teo, A. R., Chen, J. I., Kubo, H., Katsuki, R., Sato-Kasai, M., Shimokawa, N., Hayakawa, K., Umene-Nakano, W., Aikens, J. E., Kanba, S., & Kato, T. A. (2018). Development and validation of the 25-item Hikikomori Questionnaire (HQ-25). *Psychiatry and clinical neurosciences*, *72*(10), 780–788. https://doi.org/10.1111/pcn.12691

Uchida, Y., & Norasakkunkit, V. (2015). The NEET and Hikikomori spectrum: Assessing the risks and consequences of becoming culturally marginalized. *Frontiers in Psychology, 6,* 1117. https://doi.org/10.3389/fpsyg.2015.01117

Umeda, M., Kawakami, N., & World Mental Health Japan Survey Group 2002-2006 (2012). Association of childhood family environments with the risk of social withdrawal ('hikikomori') in the community population in Japan. *Psychiatry and Clinical Neurosciences*, *66*(2), 121–129. https://doi.org/10.1111/j.1440-1819.2011.02292.x

von Briel, F. (2018). The future of omnichannel retail: a four-stage Delphi study. *Technological Forecasting and Social Change, 132,* 217-229. https://doi.org/10.1016/j.techfore.2018.02.004

Wilson, S. (2010). Braindance of the hikikomori: Towards a return to speculative psychoanalysis. *Paragraph, 33,* 392–409. https://doi.org/10.3366/para.2010.0206

Wong, J. C. M., Wan, M. J. S., Kroneman, L., Kato, T. A., Lo, T. W., Wong, P. W., & Chan, G. H. (2019). Hikikomori Phenomenon in East Asia: Regional Perspectives, Challenges, and Opportunities for Social Health Agencies. *Frontiers in Psychiatry*, *10*, 512. https://doi.org/10.3389/fpsyt.2019.00512

Wong, P. W., Li, T. M., Chan, M., Law, Y. W., Chau, M., Cheng, C., Fu, K. W., Bacon-Shone, J., & Yip, P. S. (2015). The prevalence and correlates of severe social withdrawal (hikikomori) in Hong Kong: A cross-sectional telephone-based survey study. *The International Journal of Social Psychiatry*, *61*(4), 330–342. https://doi.org/10.1177/0020764014543711

Yap, S. F., Xu, Y., & Tan, L. (2021). Coping with crisis: The paradox of technology and consumer vulnerability. *International Journal of Consumer Studies, 45*(6), 1239–1257. https://doi.org/10.1111/ijcs.12724

Appendices

Appendix 1.1: PRISMA 2020 Reporting Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	10
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	11 / 83
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	14
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	15
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	16
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	16
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	86
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	16
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	17

Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	17 / 91
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	17 / 91
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	18
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	N/A
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	18
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	17
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	18
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	16

Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	16 / 17
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	18 / 19
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	21
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	27
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	26
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A

Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	32
	23b	Discuss any limitations of the evidence included in the review.	33
	23c	Discuss any limitations of the review processes used.	35
	23d	Discuss implications of the results for practice, policy, and future research.	36
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	11 / 16
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	16
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	11
Competing interests	26	Declare any competing interests of review authors.	11
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	17 / 18 / 91

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

Appendix 1.2: PRISMA 2020 for Abstracts Checklist

Section and Topic	Item #	Checklist item	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results	6	Specify the methods used to present and synthesise results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	Yes
Registration	12	Provide the register name and registration number.	Yes

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

Appendix 1.3: Systematic Review Search Strategy

Source: Embase

Interface / URL: OvidSP

Database coverage dates: 1974 to present

Search date: 17/09/2024

Retrieved records: 986

Search strategy:

1 (delphi or e-delphi).ab,sh,ti. 26909
2 "withdraw*".tw. 226649
3 exp psychosocial withdrawal/ 4951
4 social isolation.tw. 18081
5 exp social isolation/ 37270
6 2 or 3 or 4 or 5270459
7 "schizophren*".tw. 205676
8 exp schizophrenia/ 233860
9 7 or 8 263791
10 social anxiety disorder.tw. 4411
11 exp social phobia/ 15348
12 10 or 11 16080
13 exp autism/ 104904
14 pervasive development$ disorder$.tw. 2989
15 (PDD or PDDs or ASD or ASDs).tw.61995
16 autis$.tw. 93393
17 asperger$.tw. 3128
18 13 or 14 or 15 or 16 or 17 143589
19 (avoidant disorder or avoidant difficulty or avoidant personalit* or AVPD).tw. 859
20 exp avoidant personality disorder/ 1172
21 19 or 20 1701
22 depression.tw. 635292
23 exp depression/ 687501
24 22 or 23 937648
25 6 or 9 or 12 or 18 or 21 or 24 1504299

26 1 and 25 1012

27 limit 26 to english language 986

Source: MEDLINE(R) ALL

Interface / URL: OvidSP

Database coverage dates: 1946 to September 16, 2024

Search date: 17/09/2024

Retrieved records: 538

Search strategy:

1 (delphi or e-delphi).ab,sh,ti. 20242

2 "withdraw*".tw. 150784

3 exp psychosocial withdrawal/ 0

4 social isolation.tw. 13681

5 exp social isolation/ 27093

6 2 or 3 or 4 or 5185812

7 "schizophren*".tw. 144196

8 exp schizophrenia/ 118175

9 7 or 8 167193

10 social anxiety disorder.tw. 3510

11 exp social phobia/ 1357

12 10 or 11 3928

13 exp autism/ 27213

14 pervasive development$ disorder$.tw. 2112

15 (PDD or PDDs or ASD or ASDs).tw.43139

16 autis$.tw. 71879

17 asperger$.tw. 2209

18 13 or 14 or 15 or 16 or 17 89975

19 (avoidant disorder or avoidant difficulty or avoidant personalit* or AVPD).tw. 655

20 exp avoidant personality disorder/ 46193

21 19 or 20 46453

22 depression.tw. 446918

23 exp depression/ 161646

24 22 or 23 489549

25 6 or 9 or 12 or 18 or 21 or 24 927864

26 1 and 25 563

27 limit 26 to english language 538

Source: Psychology & Behavioral Sciences Collection (PBSC)

Interface / URL: EBSSCOhost

Database coverage dates: 1993 - Present

Search date: 17/09/2024

Retrieved records: 65

Search strategy:

1 AB (delphi OR e-delphi) OR TI (delphi OR e-delphi) 911

2 TI Autis* OR AB Autis* 17,413

3 TI asperger* OR AB asperger* 1,132

4 MH "child developmental disorders, pervasive" OR MH "Autistic Disorder" OR MH "Asperger Syndrome" OR MH "Pervasive Developmental Disorder-Not Otherwise Specified" 561

5 TI Avoidant Personality Disorder OR AB Avoidant Personality Disorder 159

6 MH "avoidant personality disorder" 127

7 TI depression OR AB depression 56,755

8 MH "depression" 46,236

9 TI withdraw* OR AB withdraw* 8,087

10 MH "Withdraw*" 1,950

11 TI social isolation OR AB social isolation 1,945

12 MH "social isolation" 2,372

13 TI schizophren* OR AB schizophren* 19,924

14 MH "schizophrenia" 17,657

15 TI social anxiety disorder OR AB social anxiety disorder 1,267

16 MH "social anxiety disorders" 4

17 S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 114,014

18 S1 AND S17 65

19 S1 AND S17. Narrow by Language: - English 65

Source: CINHAL

Interface / URL: EBSSCOhost

Database coverage dates: 1891 to present

Search date: 17/09/2024

Retrieved records: 259

Search strategy:

1 AB (delphi OR e-delphi) OR TI (delphi OR e-delphi) 10,341

2 TI Autis* OR AB Autis* 33,402

3 TI asperger* OR AB asperger* 1,383

4 MH "child developmental disorders, pervasive" OR MH "Autistic Disorder" OR MH "Asperger Syndrome" OR MH "Pervasive Developmental Disorder-Not Otherwise Specified" 1,641

5 TI Avoidant Personality Disorder OR AB Avoidant Personality Disorder 197

6 MH "avoidant personality disorder" 168

7 TI depression OR AB depression 154,743

8 MH "depression" 133,452

9 TI withdraw* OR AB withdraw* 30,608

10 MH "Withdraw*" 0

11 TI social isolation OR AB social isolation 7,020

12 MH "social isolation" 12,147

13 TI schizophren* OR AB schizophren* 31,403

14 MH "schizophrenia" 29,165

15 TI social anxiety disorder OR AB social anxiety disorder 1,918

16 MH "social anxiety disorders" 1,570

17 S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 310,286

18 S1 AND S17 274

19 S1 AND S17. Narrow by Language: English 274

Source: PsychINFO

Interface / URL: EBSSCOhost

Database coverage dates: 1985 to present

Search date: 17/09/2024

Retrieved records: 293

Search strategy:

1 AB (delphi or e-delphi) OR TI (delphi or e-delphi) 4,945

2 TI withdraw* OR AB withdraw* OR DE "social withdrawal" OR TI social isolation OR AB social isolation OR DE "social isolation" 65,994

3 TI schizophren* OR AB schizophren* OR DE "schizophrenia" 142,908

4 TI social anxiety disorder OR AB social anxiety disorder OR DE "social phobia" OR DE "social anxiety" OR DE "social anxiety disorder" 13,273

5 TI Autis* OR TI Asperger* OR AB Autis* OR AB Asperger* OR DE "Autism Spectrum Disorders" OR DE "Neurodevelopmental Disorders" OR DE "Autistic Traits" OR DE "Developmental Disabilities" OR DE "Neurodiversity" 90,294

6 TI avoidant personality disorder OR AB avoidant personality disorder OR DE "Avoidant Personality Disorder" 993

7 TI Depression OR AB Depression OR DE "Major Depression" 327,581

8 S2 OR S3 OR S4 OR S5 OR S6 OR S7 599,295

9 S1 AND S8 306

10 S1 AND S8. Narrow by Language: - English 293

Appendix 1.4: ACCORD Reporting Checklist

Item No	Section	Checklist Item
M1	Registration	If the study or study protocol was prospectively registered, state the registration platform and provide a link. If the exercise was not registered, this should be stated.
M2	Selection of SC and/or panellists	Describe the role(s) and areas of expertise or experience of those directing the consensus exercise.
M3		Explain the criteria for panellist inclusion and the rationale for panellist numbers. State who was responsible for panellist selection.
M4		Describe the recruitment process (how panellists were invited to participate).
M5		Describe the role of any members of the public, patients or carers in the different steps of the study.
M6	Preparatory research	Describe how information was obtained prior to generating items or other materials used during the consensus exercise.
M7		Describe any systematic literature search in detail, including the search strategy and dates of search or the citation if published already.
M8		Describe how any existing scientific evidence was summarised and if this evidence was provided to the panellists.
M9	Assessing consensus	Describe the methods used and steps taken to gather panellist input and reach consensus (for example, Delphi, RAND-UCLA, nominal group technique).
M10		Describe how each question or statement was presented and the response options. State whether panellists were able to or required to explain their responses, and whether they could propose new items.
M11		State the objective of each consensus step.
M12		State the definition of consensus (for example, number, percentage, or categorical rating, such as ‘agree’ or ‘strongly agree’) and explain the rationale for that definition.
M13		State whether items that met the prespecified definition of consensus were included in any subsequent voting rounds.
M14		For each step, describe how responses were collected, and whether responses were collected in a group setting or individually.
M15		Describe how responses were processed and/or synthesised.
M16		Describe any piloting of the study materials and/or survey instruments.

M17		If applicable, describe how feedback was provided to panellists at the end of each consensus step or meeting.
M18		State whether anonymity was planned in the study design. Explain where and to whom it was applied and what methods were used to guarantee anonymity.
M19		State if the steering committee was involved in the decisions made by the consensus panel.
M20	Participation	Describe any incentives used to encourage responses or participation in the consensus process.
M21		Describe any adaptations to make the surveys/meetings more accessible.

From: PLoS Med 21(1): e1004326. https://doi.org/10.1371/journal.pmed.1004326 For more information see: https://www.ismpp.org/accord

Appendix 2.1: Completed ACCORD Checklist for Empirical Project

Item No.	Section	Checklist Item (*help text*)	Page No.
T1	**Title**	Identify the article as reporting a consensus exercise and state the consensus methods used in the title. *For example, Delphi or nominal group technique.*	45
I1	**Introduction**	Explain why a consensus exercise was chosen over other approaches.	52
I2		State the aim of the consensus exercise, including its intended audience and geographical scope (national, regional, global).	52
I3		If the consensus exercise is an update of an existing document, state why an update is needed, and provide the citation for the original document.	N/A
M1	**Methods** Registration	If the study or study protocol was prospectively registered, state the registration platform and provide a link. If the exercise was not registered, this should be stated. *Recommended to include the date of registration.*	53
M2	Selection of SC and/or panellists	Describe the role(s) and areas of expertise or experience of those directing the consensus exercise. *For example, whether the project was led by a chair, co-chairs or a steering committee, and, if so, how they were chosen. List their names if appropriate, and whether there were any subgroups for individual steps in the process.*	53
M3		Explain the criteria for panellist inclusion and the rationale for panellist numbers. State who was responsible for panellist selection.	53-54 / 106
M4		Describe the recruitment process (how panellists were invited to participate). *Include communication/advertisement method(s) and locations, numbers of invitations sent, and whether there was centralised oversight of invitations or if panellists were asked/allowed to suggest other members of the panel.*	54

M5		Describe the role of any members of the public, patients or carers in the different steps of the study.	53-54
M6	Preparatory research	Describe how information was obtained prior to generating items or other materials used during the consensus exercise. *This might include a literature review, interviews, surveys, or another process.*	54
M7		Describe any systematic literature search in detail, including the search strategy and dates of search or the citation if published already. *Provide the details suggested by the reporting guideline PRISMA and the related PRISMA-Search extension.*	N/A
M8		Describe how any existing scientific evidence was summarised and if this evidence was provided to the panellists.	54
M9	Assessing consensus	Describe the methods used and steps taken to gather panellist input and reach consensus (for example, Delphi, RAND-UCLA, nominal group technique). *If modifications were made to the method in its original form, provide a detailed explanation of how the method was adjusted and why this was necessary for the purpose of your consensus-based study.*	54-55
M10		Describe how each question or statement was presented and the response options. State whether panellists were able to or required to explain their responses, and whether they could propose new items. *Where possible, present the questionnaire or list of statements as supplementary material.*	54 / 101
M11		State the objective of each consensus step. *A step could be a consensus meeting, a discussion or interview session, or a Delphi round.*	55
M12		State the definition of consensus (for example, number, percentage, or categorical rating, such as ‘agree’ or ‘strongly agree’) and explain the rationale for that definition.	59

M13		State whether items that met the prespecified definition of consensus were included in any subsequent voting rounds.	59
M14		For each step, describe how responses were collected, and whether responses were collected in a group setting or individually.	55
M15		Describe how responses were processed and/or synthesised. *Include qualitative analyses of free-text responses (for example, thematic, content or cluster analysis) and/or quantitative analytical methods, if used.*	57-59
M16		Describe any piloting of the study materials and/or survey instruments. *Include how many individuals piloted the study materials, the rationale for the selection of those individuals, any changes made as a result and whether their responses were used in the calculation of the final consensus. If no pilot was conducted, this should be stated.*	.54
M17		If applicable, describe how feedback was provided to panellists at the end of each consensus step or meeting. *State whether feedback was quantitative (for example, approval rates per topic/item) and/or qualitative (for example, comments, or lists of approved items), and whether it was anonymised.*	55
M18		State whether anonymity was planned in the study design. Explain where and to whom it was applied and what methods were used to guarantee anonymity.	54
M19		State if the steering committee was involved in the decisions made by the consensus panel. *For example, whether the steering committee or those managing consensus also had voting rights.*	54
M20	Participation	Describe any incentives used to encourage responses or participation in the consensus process. *For example, were invitations to participate reiterated, or were participants reimbursed for their time.*	54

M21		Describe any adaptations to make the surveys/meetings more accessible. *For example, the languages in which the surveys/meetings were conducted and whether translations or plain language summaries were available.*	54
R1	Results	State when the consensus exercise was conducted. List the date of initiation and the time taken to complete each consensus step, analysis, and any extensions or delays in the analysis.	55
R2		Explain any deviations from the study protocol, and why these were necessary. *For example, addition of panel members during the exercise, number of consensus steps, stopping criteria; report the step(s) in which this occurred.*	N/A
R3		For each step, report quantitative (number of panellists, response rate) and qualitative (relevant socio-demographics) data to describe the participating panellists.	56
R4		Report the final outcome of the consensus process as qualitative (for example, aggregated themes from comments) and/or quantitative (for example, summary statistics, score means, medians and/or ranges) data.	57-63
R5		List any items or topics that were modified or removed during the consensus process. Include why and when in the process they were modified or removed.	59 / 105
D1	Discussion	Discuss the methodological strengths and limitations of the consensus exercise. *Include factors that may have impacted the decisions (for example, response rates, representativeness of the panel, potential for feedback during consensus to bias responses, potential impact of any non-anonymised interactions).*	69-70
D2		Discuss whether the recommendations are consistent with any pre-existing literature and, if not, propose reasons why this process may have arrived at alternative conclusions.	64-69
O1		List any endorsing organisations involved and their role.	48 / 53

O2	Other information	State any potential conflicts of interests, including among those directing the consensus study and panellists. Describe how conflicts of interest were managed.	48
O3		State any funding received and the role of the funder. *Specify, for example, any funder involvement in the study concept/design, participation in the steering committee, conducting the consensus process, funding of any medical writing support. This could be disclosed in the methods or in the relevant transparency section of the manuscript. Where a funder did not play a role in the process or influence the decisions reached, this should be specified.*	48

From: PLoS Med 21(1): e1004326. https://doi.org/10.1371/journal.pmed.1004326 For more information see: https://www.ismpp.org/accord

Appendix 2.2: Approved MRP Proposal

The approved MRP Proposal can be accessed at the following link: https://osf.io/hbsd8

Appendix 2.3: Project Approval Letters

Appendix 2.3.1: Proceed to Ethics Letter

Appendix 2.3.2: Ethical Approval Letter

Appendix 2.4: Participant Documents

Appendix 2.4.1: Digital Advertisement

The digital advertisement can be accessed at the following link: https://osf.io/7prj9

Appendix 2.4.2: Confirmation of Interest and Suitability Form

The confirmation of interest and suitability form can be accessed at the following link: https://osf.io/pw6g3

Appendix 2.4.3: Participant Information Sheet

The participant information sheet can be accessed through the following link: https://osf.io/3bfxg

Appendix 2.4.4: Participant Consent Sheet

The participant consent sheet can be accessed through the following link: https://osf.io/jkex2

Appendix 2.4.5: Survey One

Survey one can be accessed by the following link: https://osf.io/4vzmt

Appendix 2.4.6: Privacy Notice

The privacy notice can be accessed by the following link: https://osf.io/km73f

Appendix 2.5: Data Analysis Plan

The Data Analysis Plan can be accessed through the following link: https://osf.io/g8hzc

Appendix 2.6: Records of Analysis

2.6.1: Quantitative Analysis

The online links to the syntax files can be found below. Each element of the survey (targets, elements, delivery, outcomes) are divided into 2-3 parts to reflect the sub-categories in the survey.

Treatment targets part 1: https://osf.io/9r7fv
Treatment targets part 2: https://osf.io/q32zy

Treatment elements part 1: https://osf.io/xhucv
Treatment elements part 2: https://osf.io/whaud

Treatment delivery part 1: https://osf.io/ynwkx
Treatment delivery part 2: https://osf.io/u278w/

Treatment outcomes part 1: https://osf.io/86dnx
Treatment outcomes part 2: https://osf.io/zyn3a
Treatment outcomes part 3: https://osf.io/s4k76

2.6.2: Qualitative Analysis

Final qualitative analysis: https://osf.io/ek2ht

Appendix 2.7: Data Availability Statement

Participants consented to their personal data being stored for the duration of the project. Following study completion, this data will be securely deleted. Anonymous research data will be kept for a minimum of 10 years, in line with the University of Glasgow policy. Participants did not consent for future data sharing; the anonymised data will support the transparency of the methodology and results only.

Appendix 2.8: Final Descriptives, Stability Statistics and Summary of Outcomes

The Final Descriptives, Stability Statistics and Summary of Outcomes spreadsheet can be accessed through the following link: https://osf.io/kfv9w

Appendix 2.9: Panel Membership Criteria

Panel	**Essential characteristics**	**Example panel members/ characteristics**	**Sample size target**
Experts by Experience	• Self-identified Hikikomori (meeting diagnostic criteria currently: _*continuous 6-month physical isolation in one's home that is significantly impacting function/causing distress; Kato et al, 2020)._ OR • Self-identified history of Hikikomori (meeting Hikikomori _diagnostic criteria_* within the last two years).	• Varied sub-types referring to severity, age, and duration. • Varied cultures. • Hikikomori with developmental disorders (e.g., autism spectrum). • Hikikomori with comorbid mood disorders. • Hikikomori that have dropped out of school/stopped working. • Recovered Hikikomori who are interested in improving care options and quality.	53
Carers of Hikikomori	• Family or carers of current hikikomori. OR • Family or carers of prior hikikomori (with a maximum of 2 years recovery).	• Varied family members (e.g., mum, dad, sibling, aunt/uncle, grandparent etc.). • Varied carers (e.g., foster parents, family friends, adoptive parents etc.).	53
Professionals	• Relevant qualification (e.g., doctorate, masters, honours, higher training, regulated health profession training). AND • Experience of work with 2+ Hikikomori and/or their families. OR • Involvement in over two publications of relevant policy documents and/or research papers.	• Clinical psychologists. • Psychiatrists. • Social workers. • Support workers. • Mental health nurses. • Academics/professors.	53
ALL	• Must be willing to engage in a Delphi study. • Must be able to self-certify that they can identify the advantages and disadvantages of participating. • Must be 18+ years of age. • Able to access the internet. • Sufficient command of English to give consent and meaningfully participate.	• Varied age ranges. • Varied cultures. • Varied duration of experience.	159