Report on a problem studied in the N-CODE Study Group April 2025

Study Group Report template

Please complete your report template and return to ncode@contacts.bham.ac.uk. Reports should be received no later than Friday 9th June and are required for the award of onward funding.

NB: This report is a summary of the activities that took place at the Study Group and the inclusion of further work on new models/solutions is not required.

Challenge - ASKSiD: Development of Large Language Model chatbot for stroke patients and carers

Report Authors - Sandeep Shirgill, Zhamayne Fakharuzi, David Cook, Timothy West, Daniel Cyrus, Inocencio Maramba

Study Group participants – Anirban Dutt , Amir Namdar, Krithika Anil, Phil Tittensor, Brendan McLean

Abstract

Stroke survivors and their caregivers often face significant barriers in accessing accurate, timely, and personalised health information, particularly after discharge from hospital care. To address this gap, the 2025 N-CODE Study Group advanced the development of ASKSiD (Ask for Stroke Information Directory), a Large Language Model (LLM)-powered virtual assistant designed to deliver tailored stroke-related information and support. Building on an initial prototype, the team worked across three integrated work packages: (1) scoping the ethical and regulatory frameworks necessary for responsible deployment of LLMs in clinical contexts, (2) developing a multi-tiered evaluation strategy to assess the quality, accuracy, and usability of ASKSiD outputs, and (3) enhancing the technological design to evolve ASKSiD from a static chatbot into a dynamic, agent-based assistant. Key achievements include a pilot scoping review on ethical deployment, early-stage human evaluation of system responses, and proof-of-concept integration with wearable health data. These efforts lay the foundation for a clinically credible, ethically robust, and user-centred digital health tool that supports proactive recovery, patient empowerment, and scalable community-based stroke care.

1 Introduction (1 page)

Stroke represents a substantial health and social care challenge in the UK, affecting over 100,000 individuals annually and resulting in approximately 1.5 million survivors (Stroke Association, 2022). Many stroke survivors experience long-term impairments, with as many as two-thirds not receiving continuous care following discharge. Recovery can extend for years, requiring sustained support beyond acute hospital settings. As outlined in Lord Darzi's report, *High Quality Care for All*, there is a pressing need to shift aspects of stroke management into community settings—improving quality of life, enabling patient autonomy, and reducing long-term healthcare costs.

A critical enabler of this shift is improved access to tailored health information. Patients and caregivers need accurate, accessible, and timely guidance to manage symptoms, navigate care pathways, and make informed decisions. However, information provision in stroke care remains fragmented. Services often operate in silos, and available resources are generic, difficult to navigate, or not responsive to the complex, evolving needs of individual users. These challenges contribute to poor health literacy, delayed interventions, and reduced confidence in self-management.

To address this gap, the N-CODE 2025 Study Group explored innovative uses of Large Language Models (LLMs) to support stroke education, building upon a platform developed with assistance of N-CODE funding from a team formed at the 2024 Study Group. This multidisciplinary team has developed and piloted *ASKSiD* (Ask for Stroke Information Directory), a virtual assistant designed to provide personalised, trustworthy stroke-related information to patients and caregivers. Built on an LLM core, ASKSiD draws from vetted sources, adapts to user needs, and is accessible via a user-friendly interface on a web browser.

The next phase of development that was initiated at the N-CODE 2025 workshop aims to move ASKSiD beyond static chatbot functionality towards an active agent based assistant by integrating features such as personalised decision support, wearable health data inputs, and action-oriented tools (e.g. calendar scheduling for therapy, medication, or lifestyle changes). These enhancements aim to deliver a scalable, ethical, and user-centred solution to improve stroke care in the community—supporting patients and caregivers not only with information but with tools to act upon it, to facilitate increased proactivity in their recovery journey.

2 Description of the problem (1 page)

Accurate, timely, and accessible information is a cornerstone of high-quality stroke care. National guidelines and research consistently show that stroke survivors and their caregivers need clear guidance to support decision-making, access services, and adjust to life after stroke (Crocker et al., 2022; Rodgers et al., 2001). Yet many report inadequate understanding of stroke, its consequences, and available support. This information gap undermines autonomy and contributes to poorer post-stroke outcomes.

Systemic barriers to effective information provision include:

1. Insufficient or overly generic information on long-term effects of stroke.
2. Poor timing—information often given too early during acute recovery when cognitive overload is high.
3. Lack of personalisation to users' literacy, impairments, or emotional readiness.
4. Weak communication between services, survivors, and carers.
5. Inappropriate formats—complex language, dense print, and limited multimodal options.

Such shortcomings have real consequences. Inadequate information can lead to unrealistic recovery expectations, disempowerment, and reduced rehabilitation engagement (Pindus et al., 2018). Many survivors describe feeling abandoned post-discharge, largely due to poor continuity of communication and fragmented community services (Aquino et al., 2020).

Large Language Models (LLMs) offer a promising solution, enabling personalised, multimodal content vital for patient education. However, their use in clinical contexts raises significant ethical, regulatory, and safety concerns. LLMs may generate "hallucinations"—plausible but incorrect information—and lack transparency, making it difficult for users or clinicians to verify outputs. The full spectrum of ethical implications of using this technology for patient education and clinical decision making is poorly understood in terms of issues such as understanding how to evaluate information delivery, how to address algorithmic biases in tailoring outputs to users' specific cognitive and cultural needs, and safeguarding of harmful or misleading content. This accompanies understanding regulatory challenges including compliance with emerging standards such as "Artificial Intelligence as a Medical Device" put forward by government bodies such as the MHRA.

Insights from the N-CODE Study Group 2025 highlight the need for a robust, multidisciplinary approach in developing the AskSID platform. We identified three key challenges:

1. Understanding the ethical and regulatory landscape of AI-powered patient education;
2. Creating methods to evaluate content quality and its impact on recovery;
3. Engineering technical solutions to maximise ethical, effective adoption.

Addressing these challenges is essential for improving continuity of care, supporting patient empowerment, and delivering equitable stroke rehabilitation. Without intervention, survivors and carers will continue to face unmet needs, fragmented support, and diminished quality of life.

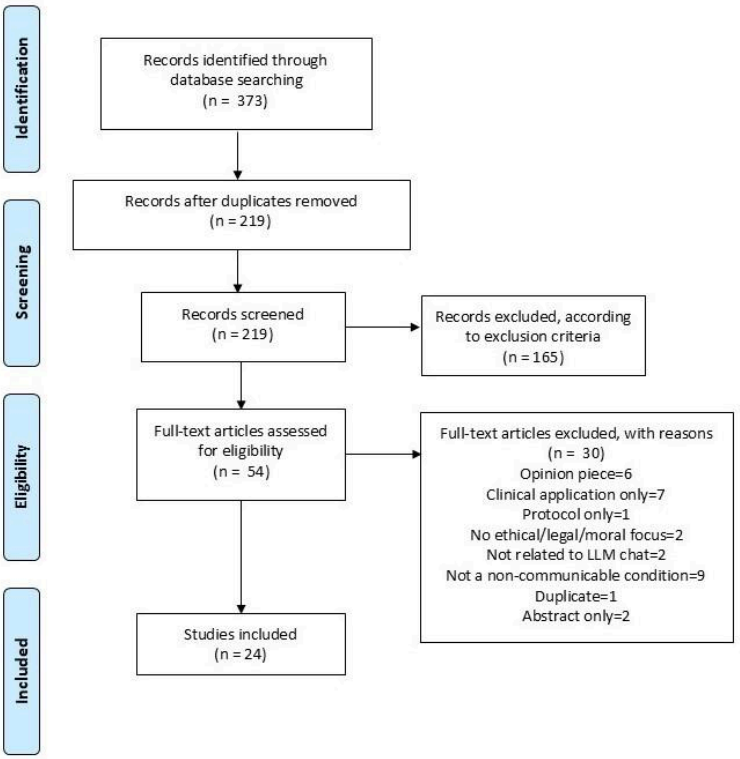
3 Solutions proposed (split into sections as required e.g. approaches considered to solve the problem, any mathematical model development and relevant results, technological solutions, etc.) (5 pages max)

During the workshop we developed the ASKSiD project within three sub-teams that related to the individuals' particular interests and skills. These three subteams were organised to provide pilot work for three interlinked work packages (WPs) that together address the ethical, evaluative, and technical challenges of delivering personalised stroke information through an AI-driven platform. WP1 will scope the ethical and regulatory landscape to ensure responsible deployment. WP2 builds on this by systematically evaluating the quality and safety of the information provided, using both human and automated assessments. Finally, WP3 leverages these insights to enhance the ASKSiD platform technologically—transitioning from a static chatbot to an adaptive, agent-based system capable of supporting personalised, action-oriented care.

Work Package 1: Neuro-AI Ethics and Regulatory Frameworks

Outline: This work package addresses the ethical, legal, and regulatory considerations essential for deploying LLMs in clinical settings. It builds on a prior scoping review initiated during the 2025 N-CODE workshop and will synthesise best practices from leading health and AI governance bodies (e.g., NICE, WHO, FDA). A co-design process involving stroke survivors, caregivers, clinicians, and health technology experts will be used to develop actionable guidance for ethical LLM deployment. Outputs will include a regulatory preparedness report, user-informed content guidelines, and an ethical framework for future translation. This work will ensure ASKSiD is both acceptable to users and aligned with emerging standards for safety, transparency, and accountability in digital health tools.

Progress Made in Workshop: Significant foundational progress was achieved during the N-CODE 2025 Study Group to support the ethical and regulatory development of ASKSiD. A rapid scoping review was initiated to explore ethical, legal, and regulatory considerations for deploying LLM-based AI tools in clinical and public health contexts. This involved a robust literature search identifying 219 sources, with 54 abstracts screened and full-text review partially undertaken. We have continued this work outside of the workshop, see the PRISMA flowchart below for the final review process.



Methodology for Rapid Scoping Review

This pilot review has already surfaced key themes—including lack of source transparency, risk of outdated information, and the trade-off between engagement and accuracy- each with direct design implications for ASKSiD. For example, it was agreed that content sourcing must be visible to users, outputs must prioritise clarity over creativity, and that default language should target a reading age of 10–12 years, with additional accessibility features such as "read aloud" functionality.

The rapid scoping review also identified a major research gap: the near absence of work examining how LLMs communicate with users affected by cognitive impairment or stroke-related language deficits. This finding underscored the need to co-design ASKSiD with these populations in mind, embedding inclusivity and accessibility into the system from the outset.

Finally, the team identified regulatory and compliance needs for future phases, including consultation with legal and AI governance experts. These findings have informed the development of an ethics and regulation steering group and will shape the co-production of an ethical deployment framework in the next project phase.

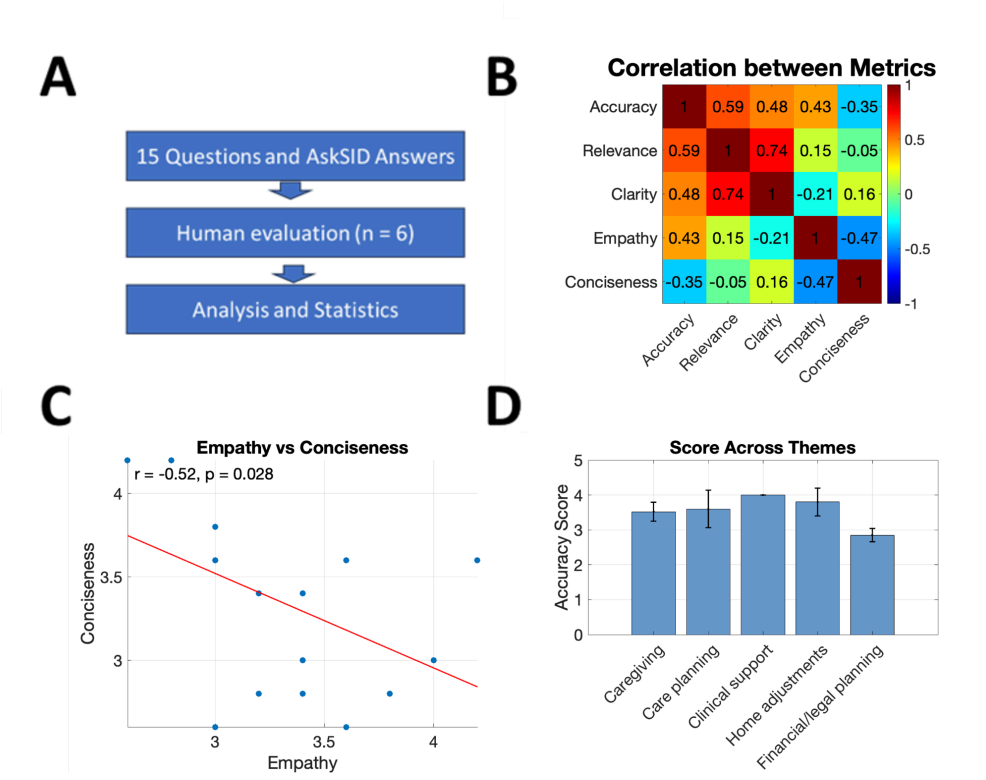
This rapid review is currently being written up for peer-reviewed, open access publication in the journal "Healthcare Technology Letters". We have also submitted an abstract for an oral presentation at the AIBA conference- international meeting based in Poland that brings together AI researchers and businesses.

Work Package 2: Evaluation of LLM Performance

Outline: Delivering safe, relevant, and user-friendly health information through Large Language Models (LLMs) requires rigorous evaluation beyond standard AI performance metrics. Work Package 2 focuses on developing and implementing a robust framework to assess the clinical accuracy, accessibility, and user experience of ASKSiD's responses. Given the high stakes of healthcare information provision, evaluations must account for both the factual integrity of the content and its comprehensibility across diverse user groups—including individuals with stroke-related cognitive impairments.

This subteam piloted quantitative and qualitative methods to evaluate ASKSiD across three key dimensions: (i) technical evaluation using expert ratings of factual accuracy and appropriateness; (ii) patient and caregiver evaluation of clarity, empathy, and usability; and (iii) planned experimental methodologies for comparative benchmarking against existing search and chatbot tools.

Progress Made in Workshop: During the workshop, an evaluation framework was developed and trialled to assess the quality of ASKSiD-generated responses. A small-scale human rating exercise was conducted amongst workshop participants (n = 6 raters) using 18 exemplar questions with responses generated by the current AskSID prototype.



Pilot Results from Work Package 3 *- **(A)** A human in the loop evaluation methodology was designed. **(B)** We analysed the relationship between rater metrics; **(C)** Initial results indicate a tradeoff between response conciseness and empathy. **(D)** Themes regarding financial and legal planning were found to lack accuracy compared to clinical questions.*

Responses were assessed across five dimensions: accuracy, relevance, clarity, empathy, and conciseness rated on a five point Likert scale. Initial findings showed high ratings for relevance and accuracy, but lower scores for empathy. A negative correlation was observed between empathy and conciseness, highlighting a trade-off between emotional tone and brevity that suggests potential for prompt optimization and personalisation to users preferences.

Preliminary results also revealed topic-specific performance variability. Responses related to financial support, legal rights, and long-term care planning tended to have lower accuracy compared to those on lifestyle and rehabilitation. These insights provide direction for targeted system improvements and can be used to guide content curation and expert review.

In parallel, planning began for a larger study involving stroke survivors and caregivers, who will rate the usability and clarity of ASKSiD responses in real-world scenarios via an online survey platform. This will inform the development of a supervised training set for a fine-tuned LLM-as-a-judge model, capable of automatically evaluating outputs at scale. The workshop confirmed the feasibility of the multi-tiered evaluation approach and helped refine key criteria and methods. These advances establish a solid foundation for scaling the evaluation process and integrating it into iterative system development.

Work Package 3: Technological Development – Agent-Based Design and Data Integration

Outline: This subteam developed a work package that aims to evolve ASKSiD from a static chatbot into a dynamic, agent-based virtual assistant capable of delivering personalised, action-oriented stroke information. Traditional chatbots typically provide a static underlying model that presents generalised responses irrespective of patients' individual needs, that is expected to limit their utility in complex care pathways where users must not only understand information but also act on it. By adopting a modular, agent-based architecture—conceptualised as a "committee of experts" - ASKSiD can deliver specialised and context-specific guidance across diverse recovery domains.

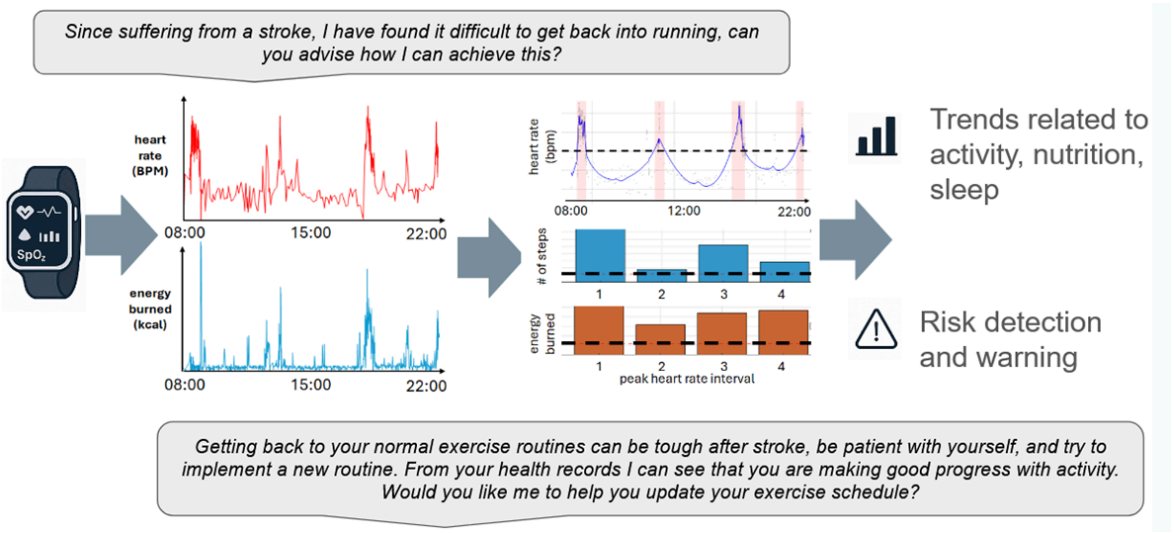
The system will comprise a set of interoperable AI agents, each responsible for a discrete function. These include:

- *Information Agent*: offering accessible, evidence-based guidance.
- *Care Navigation Agent*: assisting users in locating and accessing local services.
- *Progress Monitoring Agent*: tracking recovery metrics through integration with wearable devices and health data.
- *Action Planning Agent*: supporting users in scheduling and following through on rehabilitation goals.

This team explored the pathways for integration with user-consented data (e.g., activity levels, sleep metrics, nutrition) that can allow real-time tailoring of outputs to individual recovery trajectories. The team looked at how prompts could be engineered to tailor linguistic style, format preferences, and geographical context. Collectively, these enhancements are designed to facilitate not only better understanding, but more confident and effective user action.

Progress Made in Workshop:

The workshop supported substantial technical planning and conceptual development for this work package. A revised multi-agent architecture was outlined, with clear functional roles assigned to each agent. Prototype response flows were demonstrated to illustrate how the agents would interact, with examples including tailored advice on returning to exercise based on wearable activity data, and local service information drawn from region-specific queries (e.g., transport options to hospital appointments in Truro).



Pilot Work to Use Health Data from Wearables - *Smartwatch data was taken from volunteers at the workshop. Custom scripts were written to take data such as heart rate and energy burn to determine long range health trends, as well as stroke specific risk factors. This sets the pathway for interfacing with backend LLM prompt engineering to provide natural language advice from health data.*

Technical updates were also made to the system's interface and output features. Early feedback from workshop participants informed a change from a static chatbot to an agent based assistant with a visual avatar, as well as enhancement of the text-to-speech functionality, and prioritisation of plain-language outputs.

Exploratory work also began on integrating wearable health data. Concept prototypes were developed to demonstrate how real-time data (e.g., heart rate and activity) taken from devices such as the Apple Watch (donated by workshop participants) could be incorporated into the system to personalise feedback and support goal setting. This aligns with the long-term aim of making ASKSiD a proactive recovery companion, rather than a reactive information source.

The workshop confirmed the feasibility of the agent-based model and provided concrete design parameters for implementation. It also identified critical skills gaps—in clinical neurorehabilitation, health data integration, and regulatory compliance—which are now prioritised for recruitment and partnership development in the next phase.

4 Future work (1 page)

The next phase of the ASKSiD project will focus on systematically advancing the system from a functional prototype to a scalable, ethically robust, and clinically acceptable digital health tool. This will be achieved through development across three key domains: ethical integration, system evaluation, and technical enhancement.

Understanding the Ethical Landscape - The subteam established during N-CODE will lead co-production of an ethical and regulatory framework that will continue through stakeholder engagement with stroke survivors, caregivers, clinicians, and AI governance experts. Experts from the N-CODE network such as Dr Anirban Dutta and Phil Tittensor were keen in providing future support to ASKSiD with expert advice in regulation and clinical pathways. Outputs from this team will include user-informed guidelines for transparency, data handling, content sourcing, and safe boundaries of system functionality. A formal report on regulatory preparedness will be developed, including a mapping of ASKSiD against evolving standards for AI in health (e.g., MHRA, NICE). This work is essential to ensure ASKSiD is future proof - meeting our ambitions to scale this platform to clinical data, and wider sets of neurological disorders.

Evaluation Frameworks and "LLM as a Judge" - A full-scale evaluation of ASKSiD's outputs will be conducted based upon the pilot work performed in the workshop. This will involve structured rating of information responses by stroke survivors, carers, and health professionals. Human ratings will be used to train a fine-tuned "LLM-as-a-judge" model, enabling scalable, automated content assessment. A small-scale experimental study will compare ASKSiD with conventional search tools (e.g., Google search) to assess user comprehension, usability, and trust. These evaluations will inform iterative system refinement, particularly in relation to response tone, technical accuracy, and domain-specific reliability (e.g., financial/legal advice).

Moving from Chatbot to Agent-Based Digital Assistant- A second round of technical developments of AskSID will be pursued. The design will incorporate modular agents for care navigation, action planning, information provision, and progress monitoring. Wearable health data (e.g., step count, activity levels) will be integrated into the system for personalisation. Additional features—such as calendar tools, reminders, and adaptive outputs—will be added to enable users not only to access information but to act on it. We will also fund time for a professional front-end developer to ensure the interface is optimized for diverse ranges of users.

Overall, the planned work will transform ASKSiD from a prototype chatbot into a patient-centred virtual assistant. We are currently in early conversations with the Stroke Association to establish a co-beneficial relationship. We will also continue existing efforts to secure additional funding, and establish the commercial viability of the platform via the University of Birmingham Business development team. These developments will lay the groundwork for clinical integration, future translation to other neurological diseases, and eventual deployment with large populations alongside healthcare providers such as the NHS.

5 Discussion and conclusions (1 page)

This project addresses a persistent and widely recognised gap in stroke care: the lack of accessible, timely, and actionable information for stroke survivors and their caregivers. Despite national guidelines and service frameworks advocating for personalised information provision, current models often fail to meet user needs due to fragmented delivery, inadequate tailoring, and poor alignment with real-world contexts. AskSID proposes a novel solution that leverages the capabilities of Large Language Models (LLMs) and agent-based system design to bridge this gap, offering a scalable and adaptive digital tool that supports recovery and informed decision-making after stroke.

The N-CODE 2025 workshop has laid a strong foundation for this ambition by helping us build a strong, cohesive team and network of experts, as well as piloting our ideas. Through early-stage scoping reviews, stakeholder engagement, and technical prototyping, the project has clarified user requirements, identified key ethical and regulatory considerations, and demonstrated the feasibility of multi-agent architectures tailored to the stroke care pathway. Initial evaluations highlight promising performance in relevance and clarity, while also revealing information gaps and areas requiring novel technical solutions from the team.

The proposed future work builds on these foundations by integrating co-produced ethical safeguards, rigorous performance evaluation, and personalised system features, including integration with wearable health data and scheduling tools. This will support a shift from passive information delivery to proactive, context-aware assistance, addressing not only what patients and carers need to know, but what they need to do.

In conclusion, ASKSiD represents an important step towards intelligent, inclusive, and responsible digital support for post-stroke care. By combining technical innovation with ethical co-design and robust evaluation, the project is well positioned to deliver meaningful impact in both the short and long term—enhancing autonomy, reducing care fragmentation, and improving the everyday experience of stroke recovery.

6 References

Aquino, M. R. J. (., Mullis, R., Kreit, E., Johnson, V., Grant, J., Lim, L., Sutton, S., & Mant, J. (2020). Improving Primary Care After Stroke (IPCAS) randomised controlled trial: protocol for a multidimensional process evaluation. *BMJ Open*, *10*, Article 7. https://doi.org/10.1136/bmjopen-2020-036879

Crocker, T.F., Brown, L., Lam, N. et al., (2022) *Information provision for stroke survivors and their carers.* Cochrane Database of Systematic Reviews, 2021 (11). CD001919. ISSN 1465-1858

Pindus DM, Mullis R, Lim L, Wellwood I, Rundell AV, Abd Aziz NA, et al. (2018) Correction: Stroke survivors' and informal caregivers' experiences of primary care and community healthcare services - A systematic review and meta-ethnography. PLoS ONE 13(4): e0196185. https://doi.org/10.1371/journal.pone.0196185

Rodgers, H., Bond, S., & Curless, R. (2001). Inadequacies in the provision of information to stroke patients and their families. *Age and ageing*, *30*(2), 129–133. https://doi.org/10.1093/ageing/30.2.129

7 Appendices (For further detail if needed, e.g. any pseudo-code, mathematical analysis, etc.). can be included as a separate pdf file if needed.