Use of innovative methods in long-term follow up observational studies – an AI-enabled pragmatic literature review

Poster Code: SA100

AUTHORS: *Ian Bonzani[1], Joydeep Sarkar[2], (1) IQVIA UK Ltd and (2) IQVIA Solutions Asia Pte Ltd*

Introduction:

Long-term follow-up (LTFU) studies are a regulatory and clinical imperative for advanced therapies such as chimeric antigen receptor T-cell (CAR-T) treatments, gene therapies, and select vaccines. These modalities often induce durable biological changes—such as genomic integration or immune modulation—that may result in delayed or unexpected adverse events, leading regulators in the US and EU to mandate extended surveillance periods of up to 15 years.

The primary objectives of LTFU include monitoring for late-onset side-effects, adverse events, evaluating sustained therapeutic effectiveness, and generating real-world evidence to inform regulatory decisions and clinical practice.

However, LTFU studies face persistent challenges: patient attrition, inconsistent data capture, and logistical complexities across diverse healthcare settings. These issues will contribute to significant gaps in longitudinal data collection, limiting the ability to detect safety signals and assess long-term outcomes.

Patient engagement and retention are critical to the success of LTFU. Factors such as financial constraints and low accessibility contribute to the high rates of loss to follow-up, particularly in chronic disease populations. To address these barriers, innovative patient-centric strategies are being adopted, including decentralized study designs and the integration of digital health technologies..

This report is an AI-enabled pragmatic literature review to review LTFU studies, their use in different therapeutic areas and, to analyze the use of patient engagement tools and patient attrition rates faced in the studies.

Methods:

In this study, an AI-enabled process was used to filter publications and extra relevant data from the publications. The following step wise process was used for the selection of publications:

1. Advanced search on PubMed was used as described in **Figure 1** with a combination of keywords for search and criteria for publication period (last 5 years) and types of studies.
2. Abstracts of all 2286 papers were extracted from PubMed. The papers were divided based on the presence of PubMed Central Identifier (PMCID) and the free full text of the 880 papers with PMCID was obtained using the API shown.
3. For the use of large language model (LLM), 50 papers (abstracts, methods, results section) were used to fine-tune the detailed prompts until the prompts fine-tuned in 50% of the papers (separate prompts for abstracts, methods, results) produced a greater than 90% accuracy on the remaining 50% of the papers (**Figure 2**).

Prompts were used to identify and categorize the therapeutic area of each study (mapped to the broad categories of ICD10-CM), identify interventions (when present) which qualify as Cell & Gene Therapy (CAGT), duration of each LTFU study, use of digital technology to engage patients and collect data from patients, type of data collected through digital tools, AE data collection methods and patient payments. The number of patients at baseline and at different time points was collected to calculate patient retention metrics for LTFU studies.

Distribution of studies by disease category and duration for full text vs abstract were compared by Pearson's chi-square test and found to have no statistical difference ($p<0.05$). Hence, only full text results are shown in the figures.

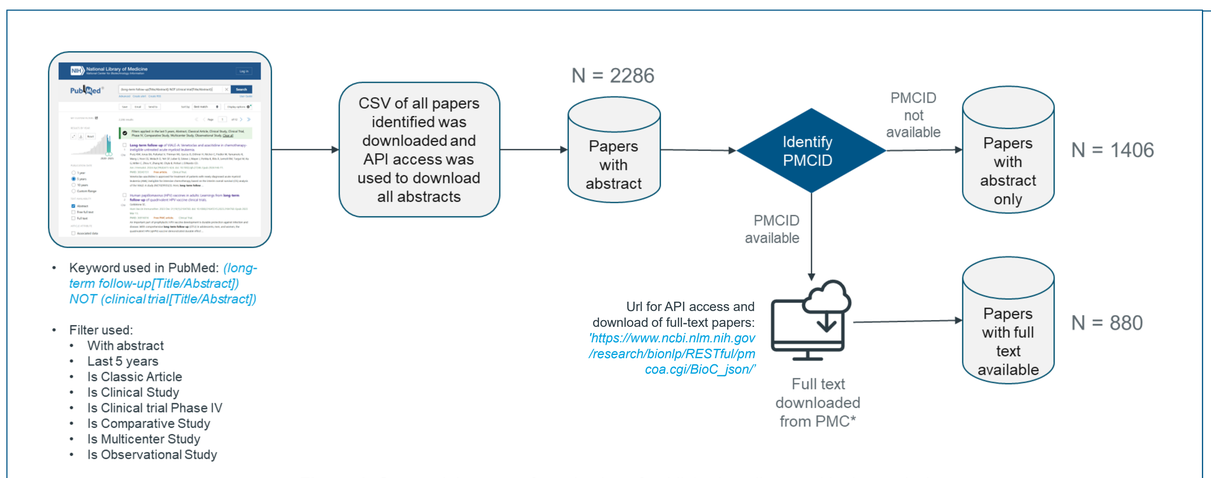

Figure 1: Sequential process for selection of published studies used in the analysis

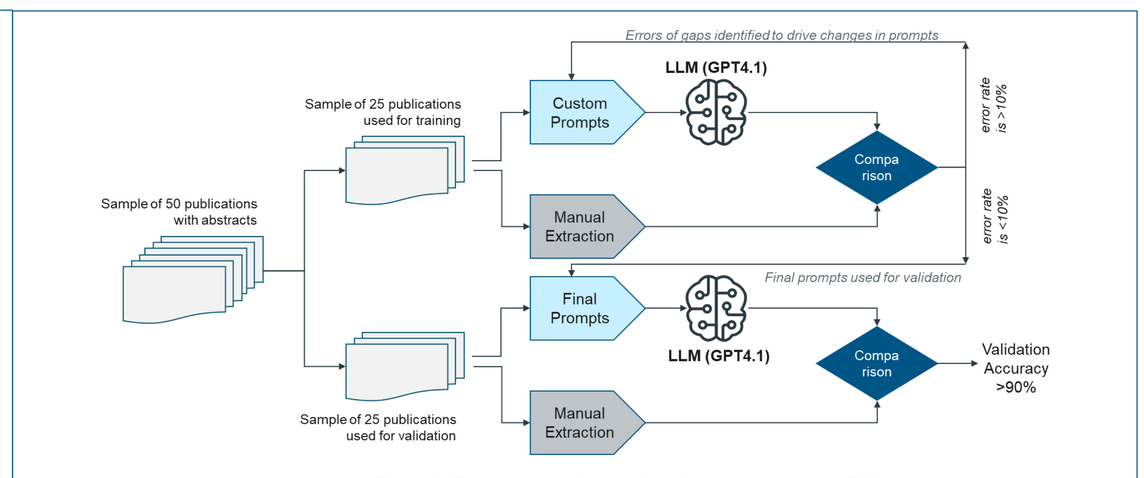

Figure 2: Training and validation of LLM prompts used in the TLR

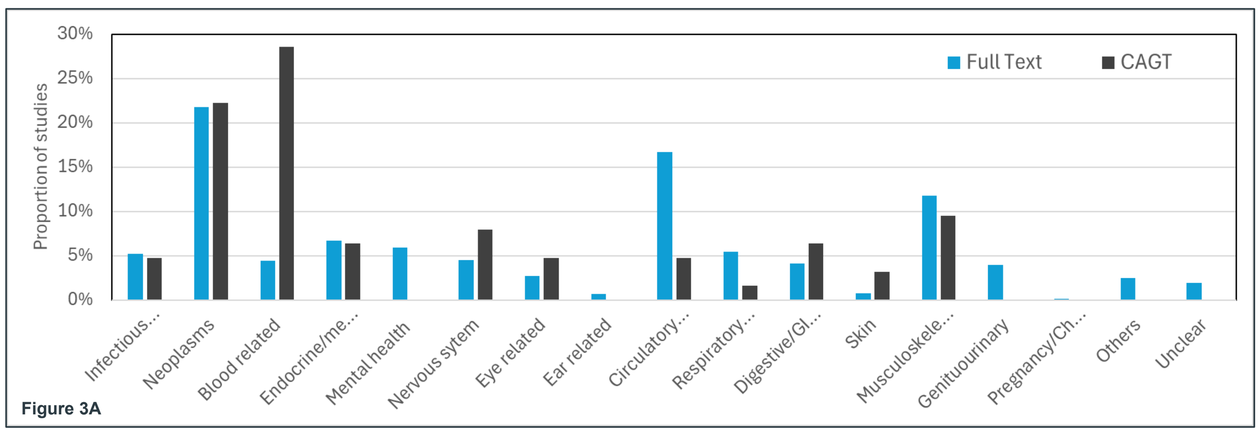

Figure 3A

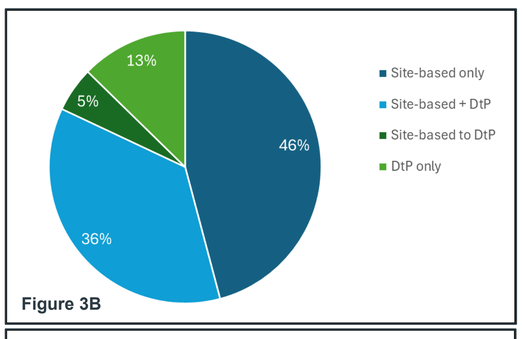

Figure 3B

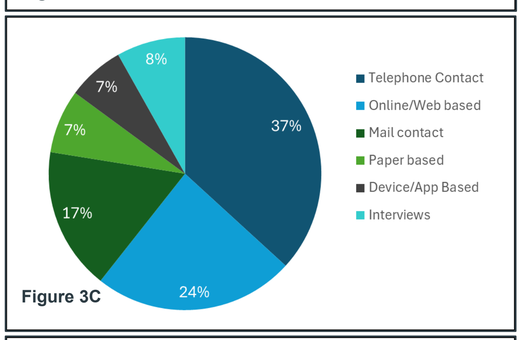

Figure 3C

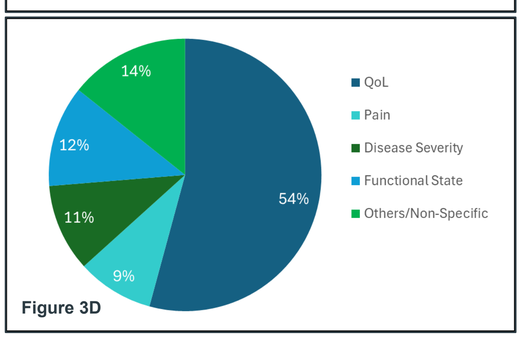

Figure 3D

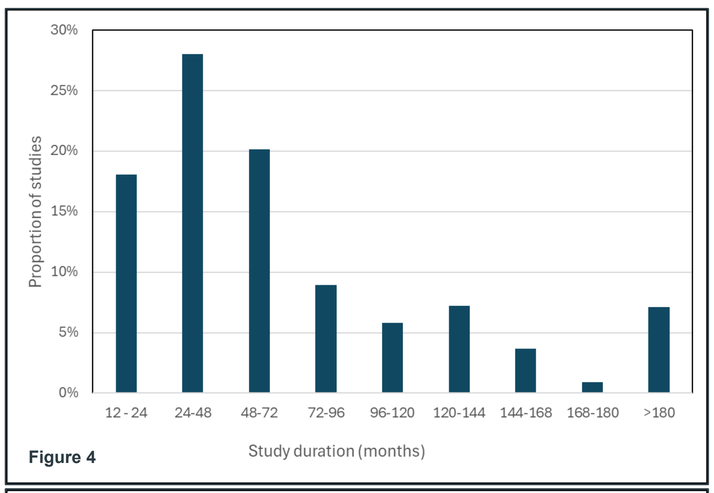

Figure 4

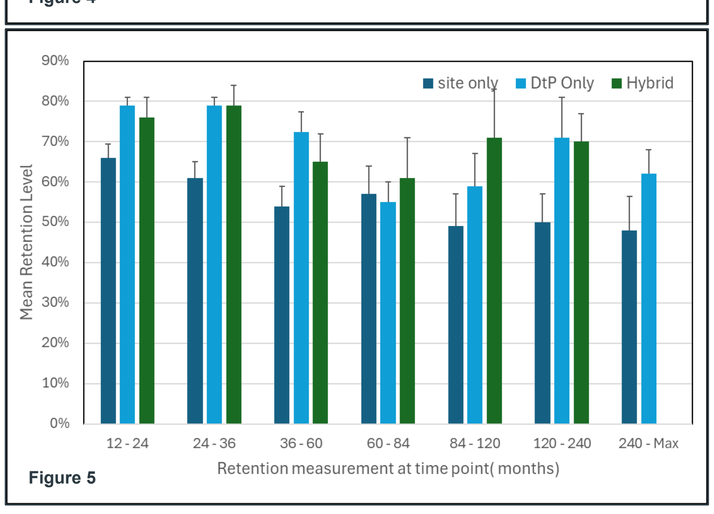

Figure 5

Results:

Figure 3A:

- Oncology related diseases had the highest number of LTFU studies, followed by circulatory/cardiovascular diseases and musculoskeletal diseases
- Majority of the CAGT studies (n=63) were for blood related diseases (CAR-T therapy)

Figure 3B:

- Majority of the studies (46%) are still using standard site-based model only
- 36% of the studies used a combination of site-based and direct to patient (DtP) methods to collect data, a small proportion of studies included (13%) used DtP only or fully decentralized approach for data collection
- We did observe that 5% of the studies transitioned from site-based to DtP approach during follow-up

Figure 3C:

- The method for collecting data directly from patients varied widely, however, 61% of studies are covered by Telephone Contact or Online/Web-based methods
- Only a small proportion (7%) used Device or App based data collection

Figure 3D:

- 55% of the studies collected adverse events (AEs) but these were mostly collected using traditional site-based methods (data not shown in figure)
- Quality of Life was the most common type of data collected from patients, with 54% of the studies collecting data from patients
- The rest was distributed between different scales or instruments to measure pain, disease severity and functional status of patients during the studies.

Figure 4:

- Majority of published LTFU studies had a duration of less than 6 years (72 months)
- A small proportion of studies (7%) had duration ≥15 years (180 months)

Figure 5:

- There is a loss of patients to follow-up as early as 1-2 years (12-24 months) for site-based studies with retention at 65% which drops to <50% beyond 7 years
- Both fully decentralized clinical studies and hybrid study models show consistently higher level of patient retention as compared to site-based studies
- The difference between site-based vs DtP or hybrid designs is particularly significant for 7-10 years follow up or longer where the difference is as high as 20% absolute

Discussion:

- This poster highlights key trends in long-term follow-up (LTFU) studies across therapeutic areas. Oncology dominates the LTFU landscape, reflecting R&D pipelines of new molecule launches and regulatory emphasis on long-term safety and efficacy. The prominence of blood-related diseases in cell and gene therapy (CAGT) studies aligns with the widespread use of CAR-T therapies, which require extended monitoring due to potential delayed adverse events.
- Despite growing use and acceptance in decentralized study approaches, 46% of studies still rely solely on site-based models. Only 13% adopted fully decentralized approaches, though 5% transitioned during follow-up, suggesting a gradual shift toward patient-centric models. Telephone and web-based forms dominate direct-to-patient (DtP) data collection, while device/app-based methods remain underutilized (7%), indicating untapped potential for digital health technology integration. Only 19 studies (out of 880) had patient payment/compensation and 10 out of them had activity-based payments.
- Quality of Life (QoL) emerged as the most frequently collected patient-reported outcome, underscoring its importance in evaluating long-term treatment impact. However, broader use of validated instruments for pain, disease severity, and functional status could enhance data granularity.
- Study duration remains skewed toward shorter timelines, with only 7% exceeding 15 years. This may reflect operational and financial constraints, as well as patient retention challenges. Retention rates are consistently higher with study designs which use tools for patient mediated data collections (hybrid & DtP) rather than site-based data collection only.
- To improve LTFU study outcomes, early integration of decentralized strategies, patient-centered technologies, and robust patient-friendly communication frameworks are essential. These approaches can mitigate attrition, enhance data quality, and align with evolving regulatory expectations for real-world evidence