PFD Toolkit: Unlocking Prevention of Future Death Reports for Research

Sam Osian [1,2,3]¶ **and Jonathan Pytches** [4]

1 Department of Primary Care and Mental Health, University of Liverpool, United Kingdom **2** Mental Health Research for Innovation Centre (M-RIC), Liverpool, United Kingdom **3** Civic Health Innovation Labs (CHIL), Liverpool, United Kingdom **4** Independent Researcher, United Kingdom ¶ Corresponding author

Summary

Prevention of Future Death (PFD) reports are used by coroners in England and Wales to alert individuals and organisations to circumstances that may lead to further deaths. Each report outlines the context, contributing factors, and concerns that, if addressed, could prevent similar tragedies.

Although these documents offer a rare, ground-level view of public safety risks, research involving PFD reports has been notoriously challenging. Assembling a usable dataset requires extensive manual effort: downloading, screening, and extracting information from thousands of inconsistently formatted documents. This has placed effective analysis out of reach for many.

PFD Toolkit is a Python package designed to overcome these barriers, automating the abundance of manual tasks involved in analysing PFD data. Using large language models (LLMs) and Vision-LLMs (supporting both proprietary and open source models), it extracts structured information from text and scanned PFD reports. Users can perform natural language searches of reports, discover recurring themes, and extract features at scale for rapid analysis.

Statement of need

Despite repeated calls for reform, PFD reports remain persistently underused in research and policy. Governmental and Parliamentary assessments have identified the current system as "under-utilised" for public safety, with no existing mechanisms to track recurring issues (Home Office, 2023; House of Commons Justice Committee, 2021). Families of the deceased have called PFD reports "nothing more than a paper exercise," highlighting the lack of systematic tracking of coroner concerns (IAPDC, 2023).

This underuse is largely driven by a series of practical and technical barriers on the Courts and Tribunals Judiciary website, where PFD reports are published: users cannot mass-download reports, formats are inconsistent, metadata is incomplete, and many reports are low-resolution scans that lack embedded text. Around 73% of reports lack thematic labels, and those that exist are inconsistently applied (Anthony et al., 2023; Zhang & Richards, 2023).

Researchers must screen and extract information from potentially thousands of reports individually, an effort that may require months or even years of manual labour (Bremner et al., 2023). One study on opioid deaths manually screened as many as 3,897 reports by hand (Dernie et al., 2023).

The *Preventable Deaths Tracker* is an existing resource that provides valuable summary statistics and metadata on PFD reports, offering the only centralised resource of its kind. However, it lacks report-text searching, thematic discovery, and custom information extraction.

⁴¹ A critical gap remains in the infrastructure for supporting scalable analysis and reducing the
⁴² manual burden of PFD research.

⁴³ Researchers have called for technology to automate data collection, enhance data quality, and
⁴⁴ surface information from unstructured text (Bremner et al., 2023; Zhang & Richards, 2023).

⁴⁵ *PFD Toolkit* addresses each of these longstanding challenges by enabling researchers to load
⁴⁶ report data; automate the discovery of recurring themes; and extract structured information at
⁴⁷ scale. This makes possible – for the first time – timely, comprehensive, and flexible analysis
⁴⁸ across the entire national archive of PFD reports.

Key features

⁵⁰ 1. **Rapid data access.** Instantly loads the latest Prevention of Future Deaths data, updated
⁵¹ weekly.
⁵² 2. **Three-layer scraping.** Handles HTML, .pdf, and scanned image reports using vision-
⁵³ enabled large language models (V-LLMs) for comprehensive data extraction.
⁵⁴ 3. **Automated text cleaning.** Corrects spelling, grammar, and formatting issues across
⁵⁵ diverse report structures.
⁵⁶ 4. **Flexible searching.** Allows users to automatically screen and filter thousands of reports
⁵⁷ using natural language queries (e.g. "medication errors", "deaths in police custody").
⁵⁸ 5. **Topic modelling.** Discovers recurring themes contained within a given selection of PFD
⁵⁹ reports.
⁶⁰ 6. **Custom feature extraction.** Pulls structured fields and variables from unstructured report
⁶¹ text.
⁶² 7. **Fast run-time.** Supports parallel processing of LLM tasks.

Example usage

⁶⁴ The example below showcases key features #1 and #4. With just a few lines of code and
⁶⁵ minutes-long runtime, users can automate what would otherwise take months of manual effort:
⁶⁶ downloading and screening thousands of reports into a research-ready dataset.

```python
from pfd_toolkit import load_reports, LLM, Screener

# -- Load reports into a pandas DataFrame --
reports = load_reports(
    start_date="2013-01-01", end_date="2025-07-01")

# -- Set up an LLM client --
llm_client = LLM(api_key=<YOUR_OPENAI_API_KEY>)

# -- Screen/filter reports by a natural language query --
query="""
Reports explicitly mentioning detention
under the Mental Health Act
"""

screener = Screener(reports=reports, llm=llm_client)

filtered_df = screener.screen_reports(
                        user_query=query)
```

Availability and documentation

PFD Toolkit is available on PyPI (via `pip install pfd_toolkit`). Full source code is available on [GitHub](#) while package documentation is available [online](#).

We welcome community contributions, feedback and feature requests. Please see our [contributions page](#) for more.

The package is unit tested with 74% coverage and is actively maintained.

References

Anthony, G., Aronson, J. K., Brittain, R., Heneghan, C., & Richards, G. C. (2023). *Preventable suicides involving medicines before the covid-19 pandemic: A systematic case series of coroners' reports in England and Wales*. https://doi.org/10.1101/2023.07.02.23292130

Bremner, B. T., Heneghan, C., Aronson, J. K., & Richards, G. C. (2023). A systematic narrative review of coroners' Prevention of Future Deaths reports (PFDs): A tool for patient safety in hospitals. *Journal of Patient Safety and Risk Management*, *28*(5), 227–236. https://doi.org/10.1177/25160435231198685

Dernie, F., France, H. S., Thomas, E. T., Bilip, M., DeVito, N. J., Ferner, R. E., Cox, A. R., Heneghan, C., Aronson, J. K., & Richards, G. C. (2023). Preventable deaths involving opioids in England and Wales, 2013–2022: A systematic case series of coroners' reports. *Journal of Public Health*, *45*(4), e656–e663. https://doi.org/10.1093/pubmed/fdad147

Home Office. (2023). *A hillsborough legacy: The government's response to bishop james jones' report (accessible)*. https://www.gov.uk/government/publications/hillsborough-disaster-report-government-response/a-hillsborough-legacy-the-governments-response-to

House of Commons Justice Committee. (2021). *The coroner service: First report of session 2021–22* (HC No. 68). House of Commons. https://committees.parliament.uk/publications/6079/documents/75085/default/

IAPDC. (2023). *'More than a paper exercise' − enhancing the impact of prevention of future death reports*. Independent Advisory Panel on Deaths in Custody. https://iapdeathsincustody.independent.gov.uk/document/more-than-a-paper-exercise-enhancing-the-impact

Zhang, Q., & Richards, G. C. (2023). Lessons from web scraping coroners' Prevention of Future Deaths reports. *Medico-Legal Journal*, *91*(3), 142–147. https://doi.org/10.1177/00258172221141284