



Blogging for students and academics: A worked example

We'll [use this paper](#) to run through a quick, worked example of how you might approach writing an effective blog post.

The title or headline

Make it catchy and informative, using active language. Consider including a number or posing a question.

Predicting Breast Cancer Mortality Risk: Insights from 11.6 Million Health Records

The introduction (1-2 paragraphs)

Start with a hook. An interesting fact, question, or anecdote. Briefly state the topic, tell people what you're going to tell them about, and why it matters. Outline what the reader will learn.

What if we could identify women at high risk of dying from breast cancer, even before they're diagnosed? Our latest research takes an important step towards making this a reality.

In this post, I'll share how our team developed a new tool that has the potential to significantly improve breast cancer screening and prevention. We'll explore our data-driven approach, key findings, and what this might mean for the future of breast cancer care in the UK.

Main body (3-5 sections)

Use subheadings to break up text. Each section should cover one main point or idea. Include evidence, examples, or data to support your points. Use short paragraphs (2-3 sentences) for readability.

Why we need better risk prediction

Current breast cancer screening in the NHS relies heavily on age, with women aged 50-70 invited for mammograms every three years. But age is an imperfect predictor. By identifying high-risk women of all ages, we could target screening and prevention efforts more effectively.

Moreover, some breast cancers are more aggressive than others. Identifying women at risk of the most life-threatening cancers could save more lives than just predicting cancer occurrence.

Harnessing the power of big data

To develop our risk prediction models, we analysed anonymised health records, cancer diagnoses, and mortality data from over 11.6 million women across England. This vast dataset from the QResearch database let us look for patterns and risk factors with unparalleled statistical power.

We used this data to develop and compare four different types of prediction models:

1. **Cox proportional hazards regression:** A statistical tool that helps researchers understand how various factors, like age or treatment type, affect the likelihood of an event occurring over time, such as a patient's risk of experiencing a heart attack.
2. **Competing risks regression:** A way to analyse situations where individuals face multiple potential outcomes that might "compete" with each other, like when a patient could experience either a stroke or heart failure and experiencing one might change the risk of the other.
3. **Extreme Gradient Boosting (XGBoost) machine learning:** A cutting-edge computer program that can "learn" from data to make accurate predictions, often outperforming other methods in complex problem-solving tasks, like predicting which patients are most likely to respond well to a particular treatment.
4. **Neural networks:** A type of artificial intelligence modelled after the human brain, capable of recognising intricate patterns in data and using that information to make decisions or predictions, like how our brains learn from experience to solve problems and navigate the world around us.

Each model aimed to predict a woman's 10-year risk of dying from breast cancer, even if she didn't have a breast cancer diagnosis at the start of that 10-year window.

Competing risks model comes out on top

After rigorously testing and validating these models work, the competing risks regression model emerged as the most promising. It showed high 'discrimination ability' (Harrell's C-index 0.932), meaning it could accurately differentiate between women who died from breast cancer and those who did not.



Importantly, the model was well-calibrated across different age and ethnic groups. This means it gave reliable risk estimates for diverse populations, which is crucial for equitable real-world use.

The model uses information readily available in GP records, such as:

- Age
- BMI
- Family history of breast cancer
- Previous hormone replacement therapy
- Medical conditions like heart disease and diabetes

This makes it practical to implement in existing healthcare systems without the need for expensive genetic tests or mammographic density measurements.

A personal perspective (optional)

Share a relevant personal anecdote or reflection to make your writing more relatable and impactful.

As a researcher who has seen the impact of breast cancer on loved ones, I'm motivated by the potential to prevent such losses. By enabling earlier detection and personalised prevention, we hope to give more women a better chance against this disease.

The Implications (1-2 paragraphs)

Discuss why this matters to readers. Include potential applications or impacts on healthcare/policy.

While our model isn't yet ready for clinical use in the real world – there's still work to be done – it does open up some promising possibilities:

- **Personalised screening:** Women identified as high-risk could be offered earlier, more frequent, or more sensitive screening, regardless of age.
- **Targeted prevention:** High-risk women could be prioritised for preventive measures like chemoprevention drugs (e.g. tamoxifen or anastrozole).
- **Expanded access:** Some younger women currently ineligible for screening could be identified as high-risk and offered screening earlier.
- **Research and trials:** The model could help recruit high-risk women into clinical trials for new screening technologies or prevention strategies.



Conclusion (1-2 paragraphs)

Summarise your key points and reinforce the main message. End with a thought-provoking statement or call to action.

Next steps on the path to impact

Our research highlights the significant potential of leveraging routinely collected health data to predict breast cancer mortality risk. By identifying high-risk women, we could move towards more personalized screening and prevention approaches that could save more lives.

But there's more work to be done before this model can change clinical practice. We need further validation in external datasets, health economic evaluations to assess cost-effectiveness, and careful consideration of potential ethical and societal implications.

Moving forward, we're excited to collaborate with patients, clinicians, policymakers and other stakeholders to refine and responsibly implement this technology. Together, we could make a real difference in the fight against breast cancer.

Read the full study in [The Lancet Digital Health](#) or visit our project pages to learn more and get involved.

About You

Add a brief bio, including your role in the department and any relevant expertise or experience. Use this to publicise yourself, share your social media handles (if desired), or link to profile pages.

Dr Jane Smith is a Senior Researcher in the Cancer Epidemiology Unit at the Nuffield Department of Primary Care Health Sciences. With over 15 years of experience in cancer research, her work focuses on using large datasets to improve cancer screening and prevention.

Follow her on Twitter @JaneSmithPhD or visit her [profile page](#) to learn more about her impactful research.