

EARLY DAYS

When you join a practice as a partner, you generally buy a share of the practice's 'fixtures and fittings'. Things like the autoclave (which gurgled alarmingly until, in one glorious explosion, was no more), and the chipped coffee mugs, plus the expensive computer. Thus, in 1990, when I joined my first practice, I was relieved that the computer came for free. *'Some company gave it to us, so they can use our prescribing data.'* The rules of this freebie — ensuring all consultations were logged, and all prescriptions had a documented indication — seemed good notekeeping practice, a rare merger of the needs of my morals and those of my wallet. A few rules were less helpful. We kept getting shirty reminders to record the Apgar score of all our births. This was a chore (and almost certainly worthless). Welcome to the world of Value Added Medical Products (VAMP) — the precursor of the General Practice Research Database, now expanded to the Clinical Practice Research Datalink (CPRD) — a government research service. The data are used in 20 countries, and research studies have resulted in more than 2000 peer-reviewed publications. Little did I know.

MODERN DAYS

Nearly 30 years later I use CPRD data almost daily. My diagnostic research — mainly in cancer — revolves around symptoms, and how they can identify possible disease. It is possible to collect data yourself, of course. In the early 2000s I had a squad of research assistants photocopying sections of patients' notes in surgeries. This included finding notes of the deceased (researching cancer has this awkward aspect). One surgery stored them in boxes in the roofspace, providing a soft landing when I headbutted a roof beam. How much better it is if someone collects the data for you — and with CPRD holding anonymised data on 35 million individual patient records by the end of 2018 (the more data, the better — CPRD is actively seeking more practices to join and contribute data,

with minimal effort on the part of practices that wish to participate),¹ it now makes it possible to study many things, such as:

- rare diseases (how else could you study the primary care features of myeloma?);
- individual symptoms (how many patients with cough would you need to collect and follow for months to capture enough with lung cancer?);
- blood test abnormalities (how common is cancer with hypercalcaemia?);
- small but important effects (do proton pump inhibitors increase pneumonia?);
- secular trends (what is the average thyroid-stimulating hormone when thyroxine treatment is started?);
- diagnostic difficulties (how much overlap is there between IBS and CFS/ME?);
- treatment effectiveness (does the flu vaccine work for older people? How long does an individual antidiabetic drug work for?);
- GP workload over time (though you can't study GP pay); and
- controversies (does the MMR vaccine cause autism?).

Of course, CPRD data are observational, needing careful interpretation and validation. This is where CPRD's evolving patient-consented studies can allow interventional studies. To date, CPRD has hosted a small number of patient-consented clinical studies — data collection is very simple, dramatically reducing the costs. This is an area with the potential to deliver significant public health benefits in the future.

WORKING DAYS

Working with the data takes a bit of getting used to (the computer equivalent of headbutting a roof beam: it's great when it's over). It comes in several giant text files via a secure download. These files are logical (consultation data, referral data, and therapy data), but contain many

ADDRESS FOR CORRESPONDENCE

Willie Hamilton

University of Exeter, Exeter EX1 2LU, UK.

Email: w.hamilton@exeter.ac.uk

unwanted variables. You remove these at the beginning, and, at the end, you realise you needed them. You can also get access to individually linked data (this is a real boon, being able to get HES data or data from the cancer registry).

The whole processing phase of the study is long — several weeks at least — and requires slickness in data manipulation. I'm now much more adept at data-handling commands in Stata than I am for statistical commands.

Finally, you can answer your research question and write the paper. Then the *BJGP* (we hope) takes it, and just like the autoclave that began this story — BOOM — the paper is on the front pages of the national press (picked up by 43 news outlets), you're on Radio 4 (who correctly give star billing to the researcher who actually does the work), and you find it's the top download in the *BJGP* for the year.^{2,3}

Willie Hamilton,

Professor of Primary Care Diagnostics, University of Exeter, Exeter.

DOI: <https://doi.org/10.3399/bjgp18X697745>

REFERENCES

1. Clinical Practice Research Datalink. General practitioner. <https://www.cprd.com/generalpractitioner/> (accessed 1 May 2018).
2. BJGP Life. Top 10 most read BJGP research articles published in 2017. <http://bjgplife.com/2018/01/30/top-10-read-bjgp-research-articles-published-2017/> (accessed 1 May 2018).
3. Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* 2017; DOI: 10.3399/bjgp17X691109.

"How much better it is if someone collects the data for you ..."