

# How and Why to Create and Maintain: A Database for Patients in Your Disease of Interest

Bruce Hough\*

Department of Internal Medicine, Virginia Commonwealth University School of Medicine, VA 23298, USA

## Abstract

One of the central requirements of a career in academic medicine is that one publishes original research while also seeing patients with that same disease state. I humbly recommend you spend a little of your protected time in the early years of your academic career starting a database in your disease of interest so you may enroll and follow these patients and potentially learn from their experience in a way that may not be obvious in clinic.

**Keywords:** Database • Mysql • PHP • Linux • Apache

## Description

One of the central requirements of a career in academic medicine is that one publishes original research while also seeing patients with that same disease state. During my years in private practice before I came to Virginia Commonwealth University (VCU) in Richmond, VA, I imagined that this required a completely different skillset and was essentially performing two jobs at once. I was pleasantly surprised when I found out just how much clinical practice and clinical/translational research complemented one another such that one really needs the other to provide the best care for the patient.

Coming to that realization is far different however than enacting that ideal. When I was tasked with seeing our aggressive lymphoma patients at my institution, my first inkling was to set up a database as there is some indication that subsets of Diffuse Large B-cell Lymphoma (DLBCL) have different prognoses and react differently to different medications. However, there was considerable resistance to this idea from several departments in our institution. The main concern seems to be protection of the patient which I was grateful I encountered. However, this protection seems to override any counterbalancing concerns for research or progression of our knowledge base. I eventually succeeding in starting and maintaining a DLBCL database at our institution with all of the requisite information and IT controls and the purpose of this paper is to guide you gentle reader in doing the same at your institution in your chosen disease state. This paper will be divided in the following chronology which appears the most logical to me: Why? Who? How?

Why? I think the best reason to have start and maintain a database in your disease state is to be fully aware of the limitations. There are dozens of large public and private databases from the NCI, to insurance companies, to next generation sequencing companies that purport to have databased that academics can access to further learning. There are also hybrids like Trinetx. However, it is often unclear how that data was acquired, in what format it is stored, how one can access it and how faithful the data is to reality. The best

approach (which is also the most laborious) is to use the primary source as your database, the EMR that your hospital system uses.

These large EMR companies are usually not happy about sharing their data. If one is able to create an alliance between the EMR companies and your bioinformatics department, you will then have to determine who is going to do the data capture and data input and how it will be viewed. There are several platforms to view data including RedCap. However, with each intermediary, there comes another separate bureaucracy. As is the case with any large system, the larger they are, the harder it is to do nimble things.

I took a different approach and applied to the IRB for an expedited review given minimal risk to subjects. This required a data management plan laying out where the data would be stored, how safe that storage was, who had access and transfer permissions, what information would be stored and how the data would be accessed (university computer only, over VPN, etc.). Fortunately, my University has a robust computer science department and they made available an automated tool to help guide me through this. There are industry guidelines for these.

Before I embarked on this, I did my best to use the tools that were purportedly already available to me. However, the public and private data sets were oftentimes difficult to access and the discrete data was generally very sparse.

Who did I approach? I approached those patients in my clinic that had the disease I was interested in studying. I spoke with them briefly about the nature of the trial and the risks (minimal) and >95% of them signed. I give them a copy of their consent and then manually input the information from their EMR into my database.

How did I do this? The computer science department allowed me a linux account with apache privileges. While the majority of physicians may not be used to using these tools, they are not hard to learn and I suspect a sizable number of younger physicians may already be comfortable with these. In short, I created a log-in page given password access only to those on the protocol. This was connected to a simple mysql backend, linux, apache and mysql are all open source software without fees to use, (but I strongly recommend you support their projects with a tax-deductible donation). I have attached the code for each below and am happy to walk any of you through this process.

I currently have 42 patients with DLBCL in our young database, but each patients has discrete data about whether they harbor a MYD88, NOTCH1, EZH2, CD79b mutation, which cell of origin they have, and whether they are CD20 negative, or CD5 positive (subsets with traditionally worse outcomes) among other things (LDH at presentation, stage, Ki-67, etc). As we learn more about the molecular underpinning of different cancers, having a list of names is invaluable to go back and query retrospectively. Inputting patients

\*Address for Correspondence: Bruce Hough, Department of Internal Medicine, Virginia Commonwealth University School of Medicine, VA 23298, USA, Tel: +8133356942 E-mail: houghb@vcu.edu; bruce.hough@vcuhealth.org

Copyright: © 2024 Hough B. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 01 October, 2024, Manuscript No. jcre-24-150593; Editor Assigned: 03 October, 2024, PreQC No. P-150593; Reviewed: 15 October, 2024, QC No. Q-150593; Revised: 21 October, 2024, Manuscript No. R-150593; Published: 28 October, 2024, DOI: 10.37421/2795-6172.2024.8.262

prospectively and then querying (with IRB approval each time) retrospectively is time-intensive in the beginning, but will likely pay rich dividends as the number in the database increases.

I humbly recommend you spend a little of your protected time in the early years of your academic career starting a database in your disease of interest so you may enrol and follow these patients and potentially learn from their experience in a way that may not be obvious in clinic. The promise of “data-mining” the large databases has not worked the way it was promised to me (perhaps others have had more success if they released their need for more granular data) without current tools and limitations [1].

---

## References

1. <https://www.ibm.com/topics/data-management-plan>

**How to cite this article:** Hough, Bruce. “How and Why to Create and Maintain: A Database for Patients in Your Disease of Interest.” *J Clin Res* 8 (2024): 262.