



User's Guide: Other General Tips

The following 10 tips developed by SGIM members John Ayanian and Ellen McCarthy succinctly summarize key issues to consider when planning and executing a secondary data analysis:

1. Start with a clear research question and hypothesis
2. Get to know your data source:
 - Why does the database exist?
 - Who reports the data?
 - What are the incentives for accurate reporting?
 - How are the data audited, if at all?
 - Can you link the data to other large databases?
3. Get good documentation of the cohort, variables, and data layout, then read the fine print
4. Consult or collaborate with researchers who have used the database
5. Line up computing resources before data arrive
6. Allow time to receive data if not publicly available
7. Learn SAS, Stata, or other statistical software so you can analyze data yourself
8. Assess outliers & missing data with plots or frequency tables
9. Consult or collaborate with a statistician on your analysis plan, especially for complex surveys with sampling weights
10. Use clinical intuition to interpret results and consult experts as needed

In addition, the following articles provide useful general advice on planning and conducting secondary data analysis.

[Observational research databases in renal disease.](#)

Shlipak M, Stehman-Breen C.

J Am Soc Nephrol. 2005 Dec;16(12):3477-84.

[Using secondary data sources for pharmacoepidemiology and outcomes research.](#)

Harpe SE.

Pharmacotherapy. 2009 Feb;29(2):138-53.