

TOP TEN LIST

TEN BEST WAYS TO DO BAD PUBLIC HEALTH RECORD LINKAGE

With apologies to David Letterman, and thanks for editorial assistance to Elizabeth Kirby and for their insights to the following Internet contributors:

Kim Hauser, University of South Florida

Phil Klein, Wisconsin Department of Workforce Development

Richard Miller, Wisconsin Bureau of Health Information

Mark Fulcomer, Pennsylvania

Kate Kvale, Wisconsin Division of Public Health

Patrick Remington, Department of Population Health, University of Wisconsin

Russel Rickard, Colorado Department of Health and Environment

Melissa Adams, University of Alabama at Birmingham

Phil Cross, NY State Congenital Malformations Registry

R.S. Kirby, December 2002

“Just have someone else do the linkage for you, then use the ‘don’t ask, don’t tell’ method perfected by the military. That way, what don’t know doesn’t hurt you!”

-- Anonymous correspondent, summer of 2002

Top Ten List: Ten Best Ways to Do Bad Public Health Record Linkage

Number 10

All for one and one for all

Always trust the Social Security Number in the database as the correct Social Security Number for that individual.

If there are duplicate Social Security Numbers for obviously different individuals (based on age, gender or other conflicting information between persons), randomly select just one. Use the latest state lottery results to obtain the random numbers.

Top Ten List: Ten Best Ways to Do Bad Public Health Record Linkage

Number 9

The Shell Game

Change the linkage identifier every time you recreate the data set. This keeps your data users guessing, plus they can't refer to specific records based on the linkage identifier. This ensures confidentiality!

If the vital statistics agency refuses to allow birth certificate numbers to be used, generate your own unique identifier based on the record's physical location in the input file. Overwrite this field each time the dataset is accessed.

Compiling the final analysis file should be a snap!

Top Ten List: Ten Best Ways to Do Bad Public Health Record Linkage

Number 8

A rose is a rose is a rose is . . .

It doesn't matter if you get twins matched correctly across files, since they are identical anyway.

If subjects share a genotype, this should entitle you to share a link.

“Ma’am, you can have any color car you want, so long as it’s black”

-- Henry Ford, 1920s

Number 7

What you get is what you see

If a variable is listed in a data dictionary it is safe to assume to you can use it for linking . . . it has always been collected, and in exactly the same manner, for the time period and geographical area related to your study.

This rule of thumb holds especially for race/ethnicity, educational attainment, and all disease, procedure, and billing fields.

Number 6

If it runs, don't fix it

Always strive to develop computer algorithms that overmatch. High match percentages are impressive and will also save staff time.

Corollary: There should never be a need to physically examine any of the source documents used in the linkage process.

Number 5

What, me worry?

Checking for duplicate records just slows down the process – this is a step that can be eliminated.

Instead, simply verify that the output dataset contains the same number of records as the largest input file. Then, proceed to conduct the analyses.

Number 4

The quality goes in, before the name

Don't bother to check for name changes. It doesn't happen often enough to change your statistics.

This is especially true for women, children who are adopted or in foster care, or the rare family that speaks Spanish or other languages, or comes from a culture where surnames are listed first.

Top Ten List: Ten Best Ways to Do Bad Public Health Record Linkage

Number 3

Only you, and you alone . . .

There is only one valid and reliable record linkage strategy – your own. Never test or evaluate it, and by all means never subject the computer algorithm to scrutiny by others!

Top Ten List: Ten Best Ways to Do Bad Public Health Record Linkage

Number 2

Black and white, or shades of gray

Deterministic linkages must be correct – after all, they are based on EXACT matches.

Why settle for a complicated probabilistic matching procedure, when you can be certain?

Number 1

Bread and Roses

Spend months of staff time discussing whether to do record linkages. Be sure to include the department attorneys, the HIPAA privacy consultant, and the division directors for each program dataset to be included. Assume that the project will take weeks to a month at most, and that once completed, next year it can be run as an overnight computer job.