

COPAFS Quarterly Meeting

June 1, 2018

Challenges for Statistical Agencies

Charles J. Rothwell

This will be a different talk than I have usually provided to COPAFS. I will not be talking about the budget, personnel, and accomplishments of NCHS, although I will be glad to respond to questions. Since this will be the last time I have a chance to talk with you, I thought it might be appropriate to examine the challenges that NCHS faces in the near future, which I think are similar in nature to those of other Federal Statistical Agencies.

It is now very close to four and a half decades since I first became involved in health statistics, which at that time was in the state health department in North Carolina. It was a watershed time for the health care sector and especially for public health for finally IT, we called it data processing back then, started to have a significant impact on our ability to monitor health status. Although a watershed time, it was a gradual tide, which rode on the crest of new technologies. I was fortunate enough to be carried along with that tide

We are faced with another watershed time... now for Federal Statistical Agencies. Bob Groves the Provost at Georgetown and former Census Director and perhaps one of our country's finest survey statisticians used this very term at a recent National Academies presentation that I attended and it was not hyperbole.

At a time when unbiased data are needed more than ever to document where we are as a nation; what has worked and what has not; and to help guide where we need to go as a nation, societal and technological changes are putting the way we collect and provide these data back to the public, decision makers and researchers at risk. Much of the problem rests with the public who volunteers to provide us this information. They have reservations about government, they do not know us, they are being bombarded by requests for information and they see failures of trusted businesses releasing their personal data

inappropriately. Perhaps of equal importance is how they consume data and information has drastically changed in just the last five years. Think of your bookstores ... how about your local library ... major newspapers losing subscribers ... how people are accessing information has changed and those choices impact how we as federal statistical agencies fit in. If we are not giving the public the information they need in the way they now access and consume information and if they do not know the data comes from us because of our lack of branding ... why should they commit time to us for our surveys? We must overcome these changes by providing them with information that they can use through their automated personal assistants, apps, bots ... we need to meet them where they are ... or we as a nation ... as a society won't have the data to govern appropriately. Our anonymity is no longer a virtue. As in voting, we need to advertise to the public why participation is important and the difference participation can make in the lives of their family, friends and communities.

At NCHS ... all our statistical systems need to be changed ... significantly changed over the next 5-6 years. The problem we face is that our statistical systems are needed in the field while these major changes take place. Why? They are the gold standard which other data collection systems both government and private adjust themselves to. That means we need funds to test ... to innovate while we continue to provide needed data now as best we can. Without these funds to innovate in a significant way, all we can do is innovate around the edges. We can keep our trains going down the track but that track is leading us ever so surely to a dead end. We should be experimenting with flying!

So let me talk about the challenges for NCHS, which are quite similar to those faced in other statistical agencies:

1. Challenge: No funding for significant innovation for statistical agencies. Without funding we will need to close down major surveys to make that change. We were lucky at NCHS when funding through

PCORI allowed us to bring a significant change in the timeliness of vital statistics. That funding will soon end.

2. Challenge: Do we always need to say something nationally before we publish and make data available? How good is good enough? In vital statistics, we are now reporting drug overdose deaths on a monthly basis for the US and for just a few states with good reporting on specific drugs used. Such incomplete and provisional reporting is highly useful for public health surveillance and encourages faster and more complete reporting by other states. If such incomplete reporting has such value, why not elsewhere? Electronic health records are our future in the reporting of health care in the US. Instead of tens of thousands of records through surveys, we could conservatively have tens of millions of records. There are though significant issues of data compatibility across vendors of these systems. Why not work with a single vendor or with a large health care conglomerate and report on findings just with their data? Such incomplete reporting could still have significant use on the treatment of rare conditions and as with vitals encourage other vendors to provide standardized data which will lead to national reporting on detail never before achieved.
3. Challenge: Computing platforms for large data sets: Whether it means analyzing millions upon millions of electronic health records or genomic data or data derived from wearables, we need larger and more flexible computing platforms than our current fixed server farms. Whether it is using a highly secure cloud environment or a large government controlled computing environment such as in Oak Ridge, we need the capability to quickly ramp up with computing power and appropriate analytic software.
4. Challenge: New staff with new skills. We need the type of staff who can take advantage of the new computing platforms while we have existing staff maintain existing systems. We especially need staff interested in using data visualization software to make more effective our quality control systems as well as with data dissemination.
5. Challenge: Response rates. Response rates continue to decline for in-person surveys. We need more research into management of

non-response bias. We need to be even more creative in the use of paradata to improve response rates. We need to look at other survey modalities to augment or supplement our surveys. We need to become known through advertising our brand and providing our data back as information to the public for their mobile devices and automated personal assistants.

6. Challenge: Re-think our surveys. How might NHANES (our physical measurement survey) be thought of as a part of NHIS (our large household interview survey)? What physical measurements could now be undertaken in the NHIS? Could NHANES measurements be reduced to an absolute minimum and done in small mobile vans or be done at fixed ~~sights~~ sites?
7. Challenge: Source data systems need integration. The timeliness and quality of reporting of drug overdose deaths would be improved drastically if we developed the capability of the systems used by our data providers to share data instead of developing another disease specific reporting system. In this case the ability of data to be passed automatically between medical examiner/coroner systems, electronic death registration systems and electronic health records would not only improve the timeliness, robustness and quality of drug overdose deaths reporting but also make us ready to report on the next public health crisis.
8. Challenge: This challenge is unique to NCHS. The National Death Index (NDI) needs to be funded through appropriations and not self-supported through charges to the researcher. Since the funds collected for the NDI go back to the states for their vital statistics systems, appropriated dollars should go to the states to support these systems so needed by the Federal government. In a similar vein, we need to consider a National Birth Index (NBI) to be able to better assess infant and child health programs. With an ~~e~~NBI, electronic health records and the NDI we would have a framework for a health registry system for the US.
9. Challenge: Data Linkage: We need to drastically improve and expand our data linkage activities with administrative data sets within HHS and across departments.

10. Challenge: Cybersecurity. : Cybersecurity is needed to be improved for all agencies including NCHS and it will be very expensive to initiate and initially I believe will unfortunately become the main stumbling block to innovation in making our data more available. In the short-run, cybersecurity will cause us in government to become even more risk adverse than we are now ... hard to believe! If we are going to make data more available if we are going to link administrative and survey data across agencies as we did when we linked HUD data with NHANES data and showed a positive impact on blood lead levels of children whose families are receiving HUD support, then we need a safe place to do that. At this point in time that safe place needs, dare I say it, not to be connected to the networked world we now work in! Thus, Research Data Centers (RDC) and the FSRDC at universities may be a good first and safe step while we develop other more flexible methods. This issue was brought up in the recent report of the Commission on Evidence-Based Policy, which argued for a National Secure Data Service.

Am I dismayed? Is this just an old man's lament about wanting things to go back to how they were? Absolutely not ... I wish I was 40 years younger with the technical tools and knowledge to take this on. We can do this, we really can! We in Federal Statistical Agencies want to change how we collect and disseminate our data. You either need to invest in us so that we can change or keep us going as we are and invest in new organizations that will do what we are doing in different modalities. Whatever is decided, an investment is needed. I am very encouraged because the President's Management Agenda has a major goal of leveraging data as a strategic asset. As I understand it, this cross departmental effort will be to provide examples of secure data sharing, developing standards for combining data, implementing the recommendations coming from the Commission on Evidence-Base Policy Making, strengthening the FSRDC's and developing common terminology which will be so necessary to future data linkage activities. Just last week I was at a conference sponsored by OMB on leveraging data as a strategic asset and I was very impressed with the energy and interest in this Administration to improve our ability to strategically

collect, use and share data across government and with the research community. This is a great time to shape the future of Federal Statistical Agencies.

Thank you!