



Accreditation of Medi-Cal, Healthy Kids
and Healthy Families Program.

Anonymity Practices in Health Care Surveys: The Dilemma of How to Protect Respondent Privacy Without Stifling Feedback From Patients

Session: 4105.0 Ethics of Privacy and Health Data

Section: APHA Ethics Special Primary Interest Group

Topic: Ethics of Privacy and Health Data

November 5, 2013

S. Rae Starr, Senior Biostatistician
Healthcare Outcomes & Analysis

Maryam Maleki, Esq.

Regulatory Affairs and Compliance

L.A. Care Health Plan, Los Angeles CA



L.A. Care
HEALTH PLAN®

Presenter Disclosures

S. Rae Starr



The following personal financial relationships with commercial interests relevant to this presentation existed during the past 12 months:

I am employed as a Senior Biostatistician at L.A. Care Health Plan – the Local Initiative Health Authority of Los Angeles County, California.

L.A. Care is a public entity competing with commercial insurers in the Medicaid and S-CHIP markets in L.A. County.

Notes:

CAHPS® is a registered trade name of the Agency for Healthcare Research and Quality (AHRQ).
HEDIS® is a registered trade name of the National Committee for Quality Assurance (NCQA).

Outline



- I. Learning Objectives.
- II. Anonymity or Actionability: Nature of the Health Survey Dilemma.
- III. Approaches to Ethically Weighing Anonymity Against Actionability.
- IV. Applied Methods for Ensuring Anonymity.
- V. Perceived Risk to Survey Respondents Might Be Overstated.
- VI. Perverse Cases: Examples of Anonymity Rules Misapplied.
- VII. Safety Valves: How the Health Sector Is Negotiating Options.
- VIII. Options For Actionability.
- IX. Recap of Learning Objectives.

I. Learning Objectives

1. Explain the reasons for anonymity in health care surveys, to protect individuals from disclosure and harm.
2. Describe the tradeoffs between anonymity and analytic rigor for improving health access for vulnerable populations.
3. Discuss impacts: types of causal analysis limited under common methods for protecting anonymity.
4. Describe qualitative and quantitative factors that increase or decrease the risk of breaching anonymity.
5. Explain how to calculate and assess the risk of breaching anonymity in survey data.
6. Describe approaches that reduce threats to anonymity in survey data.
7. Describe issues to weigh when balancing anonymity and risk to respondents in conducting analyses.



L.A. Care
HEALTH PLAN®

II. Anonymity or Actionability: The Survey Privacy Dilemma



For assessing quality of service, regulatory agencies and health plans rely heavily on opinion surveys of patients:

- Anonymity protects patients from retaliation, and protects the integrity of their assessments by preventing gaming of a survey to exclude negative respondents.
- Member demographics in anonymous survey data are important for root cause analysis of health care problems. Effectively targeting corrective actions is vital to continuous quality improvement (CQI), to conserve resources by not unnecessarily calling patients.

Balancing benefits of analytic rigor against risks to privacy:

- Health plans handle more sensitive health information (disease, sexual activity, additions, mental illnesses). But survey ratings are unique in that members are rating the health plan sponsoring the survey, creating moral hazard for health plans.
- On either side of the issue, the stakes are highest for vulnerable populations (Medicaid).

Origins of anonymity policy:

- Anonymity policies derive from *laws*, survey protocols, *research policies*, *accrediting agencies*, and *professional standards*.
- Law and regulations define ethics of human subjects protections: Nuremberg Code, Belmont Report, HIPAA; etc.
- Statistics: Reducing *bias* due to respondent *self-selection* from privacy concerns and fear of the consequences of participating non-anonymously.

Anonymity or Actionability: The Survey Privacy Dilemma (Cont.)

Anonymity is generally viewed as being protective or beneficial to the survey respondent, and that view is justified.

However, as implied in the paper's title, anonymity policy may also stifle the patient's ability to *effectively* voice concerns or complaints on surveys:



- Anonymity protects patients from retaliation, and protects the integrity of their assessments by preventing gaming of a survey to exclude negative respondents.
- Some patients actually write in identifiers on anonymous surveys (information that may get expunged without reaching the health plan): Identifying themselves, their children, their doctors, or their clinics. Written comments sometimes giving sufficient information about diseases, treatments, and clinic visits, to reveal their identities.
- Unless covered by the survey contract, text outside of the survey checkboxes, are not generally keypunched by survey firms. Hardcopy surveys are generally not returned to the health plan in anonymous or high-volume surveys, so any such feedback from patients is lost.

Indirect stifling of feedback:

- Survey cover letters generally tell patients that their responses are important to help improve the care and services that they receive. How real is that commitment real if the health plan can't drill the data down to an actionable level?
- Anonymity protects individual patients, but also protects individual perpetrators of bad service.
- Very negative responses are unusual – but those few cases are valuable. By the time a problem has harmed enough members to pose no risk to anonymity, the point of the survey may be moot. Formal complaints, grievances, and lawsuits, become a survey in themselves.

Nature of the Survey Privacy Dilemma (Cont.)



The discussion in this paper is largely about options for balancing anonymity and actionability in the release of limited datasets.

- The problem is that many common variables in combination, act as unique identifiers. (“Answer me 20 questions and I can identify anyone in the world.”)

Problems arising in defining and implementing anonymity policy:

- Rules are sometimes imported into contexts not related to the original purpose of the rule, and then spliced and mutated in borrowed combinations that weaken actionability on patient ratings and feedback.

Facets of the problem that play into the ethics of determining acceptable risks:

- Health insurers generally operate under the “business processes research” exemption in human subjects protection regulation.
- Businesses possess and analyze very sensitive data from patients.
- Survey protocols and zeal by agencies can undermine that exemption and thwart patients’ desires:
 - (First party is the health plan; second party is the patient; and the third party is the regulatory agency or the accreditation body.)
 - Third parties control funding and ratings, and make those contingent on following survey privacy protocols, layered on top of laws and regulations to punish health plans for retaliation against respondents by denying services.

Impact of Overprotection on Analytic Rigor in Quality Improvement

Types of causal analysis that are harmed or limited under common methods for protecting anonymity:

- Analysis of causes: Indicators of whether or not a patient has a given *condition or set of conditions*.
- Evaluation of programs: Indicators of the effectiveness of giving a patient a particular *treatment or set of interventions and programs*.
- The feasibility and precision of multivariate analysis is hampered by:
 - Outright prohibitions due to cell-size rules.
 - Coarse coding and aggregating continuous variables (age, months of coverage, number of visits, height and weight in fractional form);
 - Course groupings of categorical variables (ratings, demographics: ethnicity, language, geographical region, clinics and provider groups).

Cumulative nature of such restrictions:

- CAHPS surveys examine numerous topics, each of which may require different analytic variables of various types.
 - The risk to anonymity is cumulative in a given survey dataset.
 - Adding one variable eventually means discarding another.
 - Choices about which variables to include *today*, prevent adding new variables to address *tomorrow's* health problems.



III. Approaches to Balancing Risks And Rewards



Discussion of risk is appropriately focused on risks to patients. However, there are also risks to health plans for misbehavior.

- Risk should therefore be expressed as *net risk*.
- As risk to the health plan rises, risk to the patient declines.
- (The higher the amount and likelihood of penalties to the health plan, the less restrictive the anonymity policy needs to be.)

	Largely <u>Calculable</u>	Partly <u>Calculable</u>	Largely <u>Not calculable</u>
<u>For the Patient</u>			
- Risk of re-identification:	X		
- Risk of retaliation:		Boundable	
- Uncertainty of consequences for patient:			Variable
<u>For the Health Plan and/or Survey Firm</u>			
- Risk of exposure for breach of anonymity:		Boundable	
- Risk that retaliation gets discovered:		Boundable	
- Risk and size of consequences for retaliation:	X		

Rewards are somewhat more calculable:

- Benefits to patient: Exercising voice in quality of services; and the possibility that services will improve due to honest feedback.
- Reduction in harm to the patient by finding and fixing sources of bad care or services.
- Benefit to the health plan: Less E.R. use, less hospitalization, better member retention.

For a **Healthy Life**

IV. Applied Methods for Ensuring Anonymity in Health Surveys



Restriction by types of identifiers (e.g. HIPAA Safe Harbor).

- Defines risk qualitatively rather than numerically.
- Pro: Nature of risk and quality of protection are intuitive.
- Cons: Works as a one-size-fits-all rule.
- Risk is a function of the number of people in each category of the identifier. One-size-fits-all means that proper protection for the patient most at risk of being identified, will mean substantial over-protection for patients in common demographics.

Cell size rules are a common method for protecting anonymity in released data:

- Defines the minimum number of patients in the cells defined by the combination of all categories in the demographic identifiers to be released.
- Pro: Objective and calculable.
- Con: If the demographic variables for analysis are chosen poorly, the analyst is locked into that selection until the dataset is rendered obsolete over time by incoming data.

Applied Methods for Implementing a Cell Size Protocol

If demographic variables in the limited dataset are chosen poorly, the analyst is locked into that selection until the dataset is rendered obsolete by incoming data over time.



After a list of demographics is released down to the stipulated cell size, *new demographics cannot be swapped into that list* without increasing risk of breaching anonymity.

Strategies to address that limitation:

- Employ a rolling list of demographic variables in n-year sets.
- Annual surveys using random samples, can swap demographics *going forward*.
 - Reasonable to use a demographic variable for 3 years (one quality improvement cycle, and a meaningful period if pooling data over time).
- Data destruction agreements can provide a way to wipe the slate clean and start over.
 - Such an agreement may require attestation by a privacy officer and/or information security officer – someone independent of the study, with a professional stake at risk from any breach.

Cell Size Calculations

Calculating the risk to anonymity in a dataset, using the list of demographics desired in the final dataset, and the sampling frame file.

Any information known about individuals in the sampling frame by the party receiving the anonymous data, is a potential identifier.



1. List the demographics desired for release back to the health plan in the anonymous dataset.
2. Take each demographic in succession and subset it into the categories within each next demographic. This produces a tree containing all combinations of the categories in the list of demographics.
3. Count the cases in the furthest (smallest) branches of that tree, and find the smallest cell defined by that set of demographics. 1 divided by that count (for example, 10), is the risk of re-identifying a person in that cell through that list of demographics. The dataset will satisfy a cell rule of 10 or higher.

Statistical software often contains ways to automate this process for “n” demographic variables by generating “n-way” frequency tables or cross-tabulations. The counts may also be done in programming languages through a set of nested loops, with one level for each demographic variable.

Example of a Cell Size Calculation

Cell size standards are a common approach for protecting anonymity.

In Medicare CAHPS surveys, survey firms are not allowed by CMS to release cross-tabulations where any combination of demographics in the table yields fewer than 10 members in the sampling frame file.

- Yields a 10% chance of correctly guessing the identity of the respondent who gave a particular negative response on a given rating.
- A health plan would have to retaliate against all 10 individuals with those same demographics to silence the one who gave the negative rating.



Consider a random sample of $n=6$ drawn from the sampling frame below, with reporting restricted by a cell size rule of 2:

	<u>Gender</u>	<u>State</u>	<u>Rating of Health Plan</u>	
Bill	M	CA	A	"Gender" is releasable by itself because each category has at least 2 people.
Jim	M	CA	F	
Sam	M	NY	D-	"State" is releasable for the same reason.
Ted	M	CA	Not sampled	
Polly	F	NY	A+	
Susan	F	NY	A	But knowing both "State" and "Gender" would identify Sam and Jane individually, exposing them to stigma or retaliation.
Jane	F	CA	D-	
Mary	F	NY	Not sampled	

V. Perceived Risk To Survey Respondents May Be Overstated

Patients increasingly have protections and recourses against retaliation, which act as disincentives for organizations to breach anonymity and misuse survey findings to retaliate:

- Legal rights of patients in some jurisdictions.
- Appeal rights can trigger regulatory sanctions for breaches of policy.
- Possibility of legal action based on denial of care:
 - If retaliation is in the form of denying services, that could potentially be construed as fraud or abuse (taking public monies and not rendering contracted services). If so, such matters can be pursued either by the regulatory agency, or independently as a qui tam action by any party.
- News media provide the possibility of bad press and ombudsman services, alerting other customers.

Protections also exist to curb non-retaliatory nuisance uses of private information:

- Some agencies insist on prior review and approval of materials and programs contacting Medicaid members. Some also prohibit cold-calling and contacting members for recruitment or other purposes.
- As noted elsewhere, a number of survey firms subscribe to CASRO standards that prohibit using member-identified information from surveys, to individually target respondents for any purpose. When such standards are built into survey contracts, violations can entail civil liabilities for the health plan or other agency.
 - Even if a survey firm were reluctant to challenge paying customers (health plans) for violations of such standards, there is less reluctance to quietly report practices by staff to their bosses, given that both the survey firm's and the health plan's reputations are at stake.



Perceived Risk To Survey Respondents May Be Overstated (Cont.)



Retaliation against survey respondents for giving negative ratings would assume that the negative respondent will be surveyed again in an upcoming survey cycle.

- Otherwise, retaliation would serve no purpose. (Even if the retaliation is known to other to other patients, and intimidates them from giving negative responses, that same widespread knowledge of retaliation would increase the firm's risk of regulatory sanction.)
- In health plans with many subscribers, the joint probability of sampling an individual twice in separate annual surveys, is very low. For an adult population of 100,000, using the typical NCQA Health Plan CAHPS protocol, the sampling ratio is roughly 1/74.

The joint probability of being sampled twice in adjacent years, is that 1/74 ratio squared -- 0.000183 – an improbable event. Over a longer period, the likelihood that the member will remain in the health plan and hold the same negative opinion, even further diminishes the odds of a health plan benefiting from retaliation for bad ratings.

- Retaliation under a cell size rule, would require retaliating against all persons who share the same demographics. The lost capitated value of Medicaid patients in the cell (n=10) containing at least one negative responder in a low-reimbursing state might be 10 people * 12 months * roughly \$100 = \$12,000 per year lost, plus the cost of risking lawsuits, regulatory sanctions, and bad public relations. (In Medicare, the per-member monies lost in retaliating against negative respondents (if known), would be even higher.)

For a **Healthy Life**

VI. Perverse Cases: Examples of Anonymity Rules Misapplied

Borrowing privacy rules from other contexts:

- HIPAA Safe Harbor provisions are sometimes applied out of context by third parties seeking its stamp of rigorous protection. HIPAA restrictions were designed to protect specific “protected health information” (PHI), not survey data used in business processes research.



Other ways to misapply anonymity protocols, and resulting problems:

- *Less rigorous analysis:* Some public-sponsored surveys and repositories are now restricting access to *data that are already de-identified*. This may be done to limit access for competitive reasons unrelated to patient protection.
 - One practitioner in the industry noted this regarding information sharing: ~“Why is it nowadays that whenever someone doesn’t want to do something, they now cite HIPAA as the reason for not cooperating?”
- *Weaker oversight:* Stringent rules mean that health plans are unable to fully monitor vendor compliance with survey protocols.
- *Inconsistency:* Survey firms quote different cell size rules for release of limited datasets.
- *Rule saturation:* Overprotection may encourage lax compliance if it blocks the purpose of the survey or other activity.
- *Necessity unclear:* It is unknown whether respondents even believe the promise of anonymity. The limitations of anonymity may be incurred, without the desired benefits of anonymity in terms of response rates and unbiased responses,

Perverse Cases: Examples of Anonymity Misapplied (Cont.)



Risk of re-identification is largely driven by population size:

- CAHPS response datasets with some demographics are releasable to health plans. Seen through the same demographic categories, large and small health plans differ greatly in risk of re-identifying patients.
- Restriction by units such as “state” or “county” or “provider group” give capricious results:
 - Some *large health plans* have more Medicaid members than entire States in the U.S.
 - Some *counties* have more Medicaid members than entire States.
 - Some *provider groups* have more Medicaid members than entire States.
 - Some *three-digit Zip Codes* have more Medicaid members than whole health plans.
- Restrictions by data type (as opposed to cell size rules), effectively apply a more stringent rule to large health plans than to small health plans.
- This means that the larger the population in a one-size-fits-all data rubric, the lower the ability to drill down to subpopulations to identify causes of poor service quality.
 - Each respondent in a sample of 2,000 from a county of 20,000 stands for roughly 10 people.
 - Each respondent in a sample of 2,000 from a county of 200,000, stands for roughly 100 people.
 - The statistical precision of static samples may be similar, but the drilldown capability in the latter example is much lower, despite having much less risk of breaching anonymity. The result: Patients in the populous county have less likelihood of being benefited by the survey.

VIII. Options for Actionability

- Anonymity policy needs four foundations: shared understanding about what constitutes survey ethics with respect to anonymity; a rational basis (calculated risks); clarity; and uniformity.
- Build anonymity practices into policies and procedures so that practices survive staff turnover in regulatory and accrediting agencies, and regulated health plans.
- Sharing knowledge about anonymity practices may improve both the use of data, and compliance with privacy standards. At some point, seek agreement as to what constitutes norms of due diligence in anonymity practices.
- Consider a sunset policy: Census privacy rules are lifted after 75 years. (The long protection may be based on the fact that participation in the Census is compelled.)
 - Consider lifting the anonymity rule for agencies and health plans on survey ratings older than 3 years, while leaving patient privacy rules intact vis-à-vis public non-disclosure.
 - The age of the data would limit studies to perennial problems (not necessarily a bad thing), rather than single-cause events at single points in time.



Options for Actionability (Cont.)

- Inform the patient about survey risks and recourses at the time of enrollment, to strength the presumption of informed consent when patients waive anonymity on later surveys.
- Survey firms are the guardians of anonymity in many surveys:
 - Such a firm may act as a data repository, and may also run analyses on clients' behalf, presenting data only in aggregate and using a cell size rule to protect anonymity.
 - Contractual relationships change as needs evolve, or as a given firm's costs increase, or performance decreases. A change in firms may sever the health plan's access to analyses and data. This can occur in two ways:
 - The survey firm is not obliged to provide future data services without a contract, and the contracting process is often too cumbersome to justify doing so for a few analyses.
 - Outside the terms of the original contract, the survey firm may charge high fees to un-archive the data and process it.
 - The simplest solutions are in contracting:
 - Survey contracts may include survival clauses that stipulate data retention and protection after the end of the contract.
 - Similar clauses can be included that permit transfer of anonymous data to the health plan's new survey firm, if it is certified to protect anonymity.
 - Such clauses can also set costs or pricing parameters for such future services.



Options for Actionability (Cont.)

- Agency policies above the health plan level can help calibrate the balance between patient protection and data access for quality improvement. The following principles should guide policy:
 - State the rational basis (calculated risks, calculated benefits, and comparison of the risks and benefits).
 - Seek clarity (quantitative rather than qualitative).
 - Seek uniformity (apply same policy for equivalent cases or settings).
 - Define a model policy and seek to promote it through entities with bona fides in the health care industry.
 - Assemble compendia of sources on the subject matter:
 - A noteworthy example: “Report on Statistical Disclosure Limitation Methodology Statistical Policy” – Working Paper 22 (Dec. 2005) http://www.fcsm.gov/working-papers/SPWP22_rev.pdf
 - Import insights from privacy law and human subjects protection regulations.
 - Periodically recalibrate policy based on input from stakeholders (patients, health plans and other users of survey findings, consumer advocates, and ethicists). For transparency, publish the reasoning and the process as the policy evolves.



VIII. Safety Valves: How the Survey Industry Is Negotiating Options



Member-identified surveys are one solution to the restrictions:

- For Medicaid (state) and Medicare (federal), approval is sought through state or federal regulatory agencies, and the process is kept distinct from official annual surveys used by those agencies.
- Requires edits to the survey cover letters to clarify that there is no promise of anonymity, and noting what entities will and won't see the patient's responses.

Health plans sometimes have more than one patient survey in a year:

- If one survey is conducted anonymously for use in accreditation, the other survey can be member-identified.
- If the second survey is to support a pay-for-performance program, the combined sample size is good for drilldowns, as long as the samples are weighted to be representative of the health plan as a whole.

Some survey firms place restrictions on uses of member-identified survey data:

- The Council of American Survey Research Organizations (CASRO) publishes standards which, in part, oblige the client to not use the data to single out individuals for contact based on their responses on the survey. Survey contracts can cite the language of these types of standards.

Promotion of cell-size standards for release of anonymous data, is helpful in bringing clarity and objectivity to the practices of survey firms.

IX. Recap of Learning Objectives

1. Explain the reasons for anonymity in health care surveys, to protect individuals from disclosure and harm.

Health data contain very personal and sensitive information. Misuse of the information could impact coverage and benefits. Misuse of the information could harm the patient vis-à-vis outside parties (spouses, employers, customers, other insurers, etc.).

2. Describe the tradeoffs between anonymity and analytic rigor for improving health access for vulnerable populations.

For improving quality of health care and services, *targeting is everything*. Overprotecting patients can mean that corrective actions can't reach them.

3. Discuss impacts: types of causal analysis limited under common methods for protecting anonymity.

- a. Analysis of causes: Indicators of whether or not a patient has a given *condition or set of conditions*.
- b. Evaluation of programs: Indicators of the effectiveness of giving a patient a particular *treatment or intervention / program*.
- c. The precision of multivariate analysis is hampered by coarse coding and aggregating of continuous variables (age, months of coverage, number of visits, height and weight in fractional form); and coarse groupings of categorical variables (ratings, demographics: ethnicity, language, geographical region, clinics and provider groups).



Recap of Learning Objectives (Cont.)

4. Describe qualitative and quantitative factors that increase or decrease the risk of breaching anonymity.

The size of the patient population in the sampling frame reduces risk.

Sparse categories in an aggregate variable increase risk.

Missing data can act as an identifier.

5. Explain how to calculate and assess the risk of breaching anonymity in survey data.

Produce a tree reflecting the full list of demographics desired back in the dataset. Populate that tree with the cases from the sampling frame. Tally the number of persons in the furthestmost branches of that tree. Identify the smallest branch. If that branch has 11 persons, for example, 1/11 yields the risk: a 1-to-11 chance of correctly identifying one of those 11 persons through that list of demographics. The dataset would be compliant with a cell size rule of 11 or higher.



Recap of Learning Objectives (Cont.)

6. Describe approaches that reduce threats to anonymity in survey data.

Removing sparse variables is an absolute cure, but limits analysis. Aggregating sparse categories until the cell size rule is satisfied, is a common strategy.

7. Describe issues to weigh when balancing anonymity and risk to respondents in conducting analyses.

Sensitivity of the information for the respondent.

Vulnerability of the respondent to retaliation (young, infirm, mentally ill).

Degree of *incentive* present for survey clients to breach anonymity. (Rationally assess if the benefit from retaliation justify the cost and risk of retaliating?)

Degree of *disincentive* present to prevent survey clients from breaching anonymity. (Would the consequences of getting cost outweigh the cost and risk of retaliating?)



L.A. Care
HEALTH PLAN®

Contact Information

S. Rae Starr, M.Phil, M.OrgBehav, Senior Biostatistician

L.A. Care Health Plan

RStarr@LACare.org, rae_starr@hotmail.com

213-694-1250 x-4190

Maryam Maleki, Esq.

L.A. Care Health Plan

MMaleki@LACare.org

213-694-1250



L.A. Care
HEALTH PLAN®