INFERRING HEALTH INTERVIEW RESPONSE BIAS FROM IMPERFECT RECORD CHECKS*

Kent H. Marquis, The Rand Corporation

Record checks are used to infer something about survey response bias. But they can produce misleading conclusions. This paper has four objectives:

1. To show the difference between a model-based definition of response bias and the underreport-overreport definitions used in most health survey validity studies,

2. To show that, for survey reports of hospital episodes, the type of record check design has, in practice, determined the conclusions about survey bias,

3. To show that a currently-fashionable record check design modification didn't overcome design problems in one instance and,

4. To illustrate that some of our ideas about proxy and memory bias, based on record check studies of hospital episode reporting, should be reexamined.

TYPES OF DESIGNS

Three kinds of record check designs are evident from Exhibit 1.

An AB design obtains estimates for cells A and B by conducting the survey first and then checking records for sample elements within "yes" survey values.

An AC design starts with "yes" record values and interviews those sample elements.

There are several ways of conducting the full design. Their common characteristic is the production of unbiased, empirical estimates of all four cells, or at least cells A, B, and C.

DEFINITIONS

The 3 matrices below in Exhibit 2 are the basis of the model. At the left is a table showing survey response probabilities conditional on the true values. The next matrix contains record value probabilities also conditioned by the true values. The third matrix contains the true proportions of yes and no answers in the population.

EXHIBIT 2 - RESPONSE MODEL FOR A DICHTOMOUS VARIABLE

Multiplying these three matrices and assuming that record and survey errors are independent, yields the matrix (Exhibit 3) which contains the expected proportions of each kind of survey and record cross-classification for the population.

EXHIBIT 3 - EXPECTED PROPORTIONS OF CROSS-CLASSIFIED OBSERVATIONS ASSUMING INDEPENDENCE OF SURVEY AND RECORD ERRORS

Design Types:

AB (or prospective, ex ante, follow-back) checks records if the survey value is "yes".

AC (retrospective, ex post) interviews persons if the record value is "yes".

FULL obtains unbiased empirical estimates of all cells or at least cells A, B, and C.

-----

*This research was supported by the Health Insurance Study, a grant from the U.S. Department of Health, Education, and Welfare to The Rand Corporation. The opinions and conclusions expressed herein are solely those of the author, and should not be construed as representing the opinions or policy of any agency of the United States Government.

265
Using this terminology it is possible to list various definitions of survey bias that have been used in record check studies and the definition of bias from the model.

DEFINITIONS OF SURVEY BIAS:

TRUE SURVEY BIAS = \( S_p - T_p \)

\[ AB \text{ DESIGN} = S_{TP}R_{FN} + S_{FP}R_{TN} \]

\[ AC \text{ DESIGN} = -(S_{TN}R_{FP} + S_{FN}R_{TP}) \]

\[ FULL = S_{TP}R_{FN} + S_{FP}R_{TN} - S_{TN}R_{FP} - S_{FN}R_{TP} \]

In the appendix, I derive the expressions for the deviation between the design definitions and the true survey bias for the dichotomous case. Elsewhere (Marquis, 1978) I go through a similar exercise for the multinomial case.

MODEL BASED RESULTS

Some of the main conclusions from this exercise for the dichotomous variable case are:

1. That the AB and AC definitions of bias include half the random error in the survey and half the random error in records (and due to matching); in other words, they tend to exaggerate the bias estimate.
2. That the direction of the interpreted survey bias is predetermined when the AC or AB design is used.
3. All the designs mistake record bias for survey bias in their interpretations.

The record check design effects are similar when the variable is multinomial but there is less certainty about the predetermined sign of the survey bias and potentially less of the random error in the cross-classified observations is interpreted as net survey bias.

EMPIRICAL STUDIES OF HOSPITAL STAY RESPONSE BIAS

Exhibit 4 shows the link between this conceptual discussion and practice. It contains the results of a number of record check studies of hospital stay reporting in surveys. As predicted, the AC design studies find underreporting to be greater than overreporting; the AB design studies find overreporting to be greater than underreporting. Note that the ABC design studies find overreporting and underreporting of about the same magnitude. Employing some assumptions, this suggests that the cross-classified observations contain mostly, or only, random error. The one directional studies detect one part of random error and run the risk of concluding that it is survey bias.

MENTIONED PROVIDER VERIFICATION DESIGN

Contemporary health survey practice is to use the AB record check design and also to contact other sources of health care mentioned in the interview even if the source was not used by the sample person. My question is, "Does incorporating the extra feature overcome the intrinsic faults of the pure AB design?"

<table>
<thead>
<tr>
<th>Design Type and Reference</th>
<th>Percent Underreport</th>
<th>Percent Overreport</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC Designs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balcun (1954)</td>
<td>-14</td>
<td>11</td>
</tr>
<tr>
<td>Feather (1972)</td>
<td>-11</td>
<td>14</td>
</tr>
<tr>
<td>AC Designs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balcun et al. (1961)</td>
<td>-13</td>
<td>---</td>
</tr>
<tr>
<td>Cannell et al. (1961)</td>
<td>-13</td>
<td>3</td>
</tr>
<tr>
<td>Cannell and Fowler (1963)</td>
<td>-17</td>
<td>4</td>
</tr>
<tr>
<td>Kirchner et al. (1969)</td>
<td>-13</td>
<td>---</td>
</tr>
<tr>
<td>AB Designs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andersen and Anderson (1967)</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>Balcun et al. (1961)</td>
<td>---</td>
<td>3</td>
</tr>
<tr>
<td>Barlow et al. (1960)</td>
<td>-5</td>
<td>13</td>
</tr>
<tr>
<td>Barlow et al. (1960)</td>
<td>-1</td>
<td>4 to 8</td>
</tr>
<tr>
<td>Kirchner et al. (1969)</td>
<td>---</td>
<td>15</td>
</tr>
<tr>
<td>Loewenstein (1969)</td>
<td>---</td>
<td>12</td>
</tr>
</tbody>
</table>

NOTES: a. All studies exclude stays shorter than overnight and report estimates for a 12-month reference period. For an explanation of the studies and the estimates, see Marquis (1978). b. MED-10 study. c. Procedure A. d. Hospital check. e. Prospective check. f. Blue Cross check. g. Hospital check. h. Prospective check. i. Prospective hospital check.

The Rand Health Insurance Study conducted some AB design record checks of interview data and contacted other mentioned providers also. Survey bias was estimated in two ways: by comparing the survey estimate of the mean to an independent, external estimate and by more conventional record check approaches. An overview of the hospitalization results is in Exhibit 3.

The interviews were conducted in Dayton, Ohio in the middle of 1974 using a 12-month retrospective reference period.

The survey estimate of the per-person, hospital-stay mean is somewhat lower than a synthetic estimate based on the Hospital Discharge Surveys. Part of the difference is caused by using a retrospective, 12-month reference period in the survey. Assuming minimal sampling, non-response, processing and interviewer biases, and that the HDS estimate is correct, a full-design record check should reveal a small, negative response bias in the survey. The record check, however, did just the opposite.

The pure AB design record check shows no survey overreports and the incorporation of the mentioned provider feature (which checked records for an additional 900 person-hospital pairs) did not affect the pattern importantly. It resulted in only 20 more apparent survey underreports. I conclude that, in this one example, the AB design, coupled with a mentioned provider verification, was unable to obtain an unbiased estimate of the survey bias. This case study is offered to show that a commonly used modification to the AB record check design does not necessarily remedy its basic deficiencies.
EXHIBIT 5A -- ESTIMATES OF THE ANNUAL HOSPITAL STAY MEAN PER PERSON

<table>
<thead>
<tr>
<th>Estimate Source</th>
<th>Number of Hospital Stays Per Person Per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dayton survey</td>
<td>.134</td>
</tr>
<tr>
<td>Correction for retrospective sampling bias</td>
<td>Up to .008</td>
</tr>
<tr>
<td>Hospital Discharge Record Survey(^{a})</td>
<td>.16</td>
</tr>
</tbody>
</table>


EXHIBIT 5B -- RECORD CHECK ESTIMATE OF RESPONSE ERRORS

<table>
<thead>
<tr>
<th>Record Survey</th>
<th>Percent Overreport... 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Percent Underreport:</td>
</tr>
<tr>
<td></td>
<td>Pure AB design........ 02</td>
</tr>
<tr>
<td></td>
<td>with modification... 07</td>
</tr>
<tr>
<td>No</td>
<td>+20</td>
</tr>
</tbody>
</table>

| Potential Conclusion: Large positive survey response bias |

CONCLUSION OF CORRELATES OF RESPONSE BIAS
- Record check data suggest that proxy reports of hospitalization are more biased than self-reports and that memory decay is a dominant feature of recall-based reporting. However, variables correlated with the amount of random error may appear as correlates of bias using a one-directional record check design. A reexamination of available published data suggests we may wish to temper these conclusions somewhat.

Several AC record check studies (Cannell et al., 1961; Cannell and Fowler, 1963) suggest that hospitalization underreporting by proxy is greater than underreporting by self-respondents. A single AB design study (Barlow et al., 1960) finds that there is more overreporting by proxy. Two studies employing experimental variation in the respondent rule, but not record checks (Kovar and Wright, 1973; Enterline and Capt, 1959), find no significant difference between treatments in numbers of episodes reported. A reinterpretation is presented elsewhere in greater detail and suggests proxy responses contain more random response error than self responses, that self responses contain little or no net bias, and that proxy responses contain a smaller net bias that previously claimed. A full design record check study with random allocation of respondent rule treatments would be very useful in pinning down the remaining uncertainties.

The memory decay hypothesis is one of the main cornerstones of applied survey measurement design. Several AC design, record check studies support the decay hypothesis for hospitalization reporting.\(^{10}\) One published study (Feather, 1972) estimates both under- and over-reports for 3 recall period lengths. These data (Exhibit 6) suggest the absence of a memory decay bias for most recent or only hospital episodes\(^{11}\) because the number of survey-reported episodes in the most distant past is not less than the number identified in the records. The survey bias, if any, appears to be a net overreporting of hospital episodes in the near past. The underreport rates follow the same pattern published by previous AC design studies; but, because both over and underreporting increase over time without an increase in net bias, we can infer that the random error in the cross-classified observations is really what increases with elapsed time and that it is this increasing random error that the previous one-directional record check studies have mistakenly interpreted as memory decay bias.

EXHIBIT 6 -- EFFECT OF RECALL INTERVAL ON HOSPITAL STAY REPORTING IN A FULL DESIGN RECORD CHECK STUDY\(^{a}\)

<table>
<thead>
<tr>
<th>Recall Interval(^{b}) (Wekes)</th>
<th>Number of Hospital Episodes(^{c}) (12 months)</th>
<th>Percent Percent Under-report Over-report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-18</td>
<td>173</td>
<td>174</td>
</tr>
<tr>
<td>19-36</td>
<td>157</td>
<td>159</td>
</tr>
<tr>
<td>37-52</td>
<td>121</td>
<td>123</td>
</tr>
<tr>
<td>Not Reported</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

\(^{a}\)Adapted from Feather (1972). The recall interval classification is determined by the record value unless the survey report could not be matched to a record report.

\(^{b}\)Elapsed time between the discharge and the survey.

\(^{c}\)Most recent and only episodes

CONCLUSIONS
- In summary, if a record check is used to estimate survey bias, the design should include a provision to separate the random error in the cross-classified observations from the bias. Without this feature, one-directional designs run a risk of overstating the bias in a direction predetermined by the kind of design used.

Today’s survey measurement designs have become very complex and expensive, especially in the health field. Much of the complexity appears to be in response to findings of one-directional record checks. The use of self-respondents, panel designs with short recall periods, double-sampling schemes, record verifications and bounded recall procedures often can be traced to studies using incomplete record check designs. It may be cost-effective to take another look at these studies or replicate them using more complete record check designs. The possibility exists that some kinds of survey responses are less biased than
our methodology studies suggest, and that the
data collection design and data analysis modifi-
cations we've made could be doing more harm than
good.

FOOTNOTES

A third source of error, produced by matching-
and processing mistakes, is not shown ex-
citly in Exhibit 3. These errors have the same
effects on interpretation of bias as record mis-
classifications.

2Using notation in Exhibit 1, the overreport
rate is b/(a+b); and the underreport rate is
a/(a+c). These are the bias estimates from each
design relative to a mean or proportion. Using
notation in Exhibit 3, a+b = 5p and a+c = 5h.
The circumflex denotes that these are estimates
based on samples.

3The overreport and underreport rates do not
have the same denominator (see previous footnote).
However, the interpretation of the sum of the
two rates as proportional to the estimated net
bias is not sensitive to small differences be-
tween the denominators.

4Based on a stratified, random sample of
2,000 families excluding families headed by per-
sons over 64 years of age. The response rate was
80 percent of eligible families. Adults (over 17
years) responded for themselves. Data were
obtained for approximately 4,400 persons. Exhib-
it 3 is based on data from 4,256 persons who
were under 65 years of age.

5See Simmons (1967) for a discussion of the
retrospective sampling bias due to excluding
hospitalized persons who have died in the past
12 months from the survey sampling frame. See
Marquis (1978, pp. 57-63) for additional elements
of the retrospective sampling bias.

6Forms listing the name of the sample per-
son were mailed to hospitals with a request that
the hospital furnish information about each stay
for the person. 127 stays reported in the survey
for persons under 65 were excluded from the anal-
ysis as follows: 86 hospital nonresponse, 2 hos-
pital refusals, 19 family refusals to allow a
record check, 20 for other reasons.

The hospital was reported to be the sample per-
son's usual source of outpatient care or it
provided outpatient care for the sample person
or any member of his family in the past 12 months.

8Other interpretive possibilities that ap-
pear very improbable are that the synthetic HDS
estimate seriously overstates the Dayton mean or
that there are large, negative biases due to
missing observations and/or that the survey
sample is a biased population sample.

9The conclusion reached above may not gen-
eralize to other survey variables. For example,
I have reported elsewhere (Marquis et al., 1976)
the results of the dental expenditure record
check on this same survey sample using the modi-
fied AB design. There neither the comparative
means nor the record check indicated an impor-
tant survey bias. My hunch is that the dental
mentioned provider feature included most of the
universe of dentists likely to have treated
each sample person whereas the hospital men-
tioned provider feature did not. I have no indepen-
dent support for this hypothesis.

REFERENCES

Andersen, Ronald, and Odin Anderson, A Decade of
Health Services, University of Chicago Press,
Chicago, 1967.

Balamuth, Eve, Sam Shapiro, and Paul M. Densen,
"Health Interview Responses Compared with Medi-
5, June 1961, U.S. Public Health Service Publi-
cation No. 584-D5; reprinted in Vital and Health

Barlow, Robin, James Morgan, and Grover Wirick,
"A Study of Validity in Reporting Medical Care
in Michigan," Proceedings of the Social Statis-
tics Section, American Statistical Association,

Bell, Barbara B., "Validation of Morbidity Survey
Data by Comparison with Hospital Records,"
Journal of the American Statistical Association,

Cannell, Charles F., and Floyd J. Fowler, "Com-
parison of Hospitalization Reporting in Three
Survey Procedures," Health Statistics, Ser. D
No. 8, January 1963; reprinted in Vital and
Health Statistics, Ser. 2, No. 8, July 1965.

Cannell, Charles F., Gordon Fisher, and Thomas
Bakker, "Reporting of Hospitalization in the
Health Interview Survey," Health Statistics,
Ser. D, No. 4, 1961; reprinted in Vital and
Health Statistics, Ser. 2, No. 6, July 1965.

Croner, Charles M., The Nation's Use of Health
(HRA) 77-1240, 1977.

Dent, Philip E., and Katherine G. Capt.,
"A Validation of Information Provided by House-
hold Respondents in Health Surveys," American
Journal of Public Health, Vol. 49, 1959,
pp. 208-212.

Feather, Joan, A Response/Record Discrepancy
Study, University of Saskatchewan, Saskatoon,
November 1972. (Available from the National
Technical Information Service, Springfield,
Virginia.)

Kirchner, Corinne, Raymond C. Lerner, and Odette
Clavery, "The Reported Use of Medical Care
Sources by Low-Income Inpatients and Out-
patients," Public Health Reports, Vol. 84, 1969,
pp. 107-117.

Kovar, Mary Grace, and Robert A. Wright, "An Ex-
periment with Alternate Respondent Rules in the
National Health Survey," Proceedings of the
Social Statistics Section, American Statistical
Association, 1973, pp. 311-316.

Lewis, W. Frank, "Utilization of Short-Stay
Hospitals: Summary of Nonmedical Statistics,
United States--1973.1 Vital and Health Statis-
tics, Series 13, No. 23, July 1976.

Marquis, Kent H., Record Check Validity of Survey
Responses: A Reassessment of Bias in Reports
R-2319-HEW, May 1978.

Marquis, Kent H., M. Susan Marquis, and Joseph
P. Newhouse, "The Measurement of Expenditures
for Outpatient Physician and Dental Services:
Methodological Findings from the Health Insur-
ance Study," Medical Care, Vol. 14, 1976,
p. 913-931.
**APPENDIX**

This appendix derives the expression of the survey bias definition for each record check design in terms of true survey bias, true record bias, and the random error in each source. The expressions are summarized in Exhibit 8.

True Survey Bias, \( SB \), is defined as the difference between the proportion of positive survey answers, \( Sp \), and the population proportion of true positive attributes, \( Tp \). Using the definitions in Exhibits 2 and 3:

\[
SB = TP(1 - SP) + FN(1 - Tp) - Tp
\]

True Record Bias, \( RB \), is similarly defined and may be reduced to:

\[
RB = FN(1 - SP) + FP(1 - Tp)
\]

Joint errors, \( JE \), are misclassifications that occur in both the survey and record sources:

\[
JE = FP FN + FN TP
\]

Next, the expressions for the 3 design definitions of survey bias in terms of the above definitions can be derived.

For the full design:

\[
SB(ABC) = Sp - Rp = Tp SP + FN TP - Tp SP
\]

Substituting \( TP = (1 - SFp) \), \( TN = (1 - SFp) \)

and analogous record terms,

\[
SB(ABC) = SP FN - FP TN - (TP SP + TN SP)
\]

For the AB design:

\[
SB(AB) = SP FN + Tp SP - Tp SP
\]

Using \( TP = (1 - SFp) \) and \( TN = (1 - SP) \),

\[
SB(AB) = FP TN + SP TN - (TP SP + FN SP)
\]

Adding and subtracting \( SP FN + FN TP \) and \( RP FN \),

\[
SB(AB) = (SP FN - SP FN) + (FP TN - RP FN) + (RP FN - SP FN)
\]

Using the definitions at the beginning of the appendix,

\[
SB(AB) = SB + (SP FN) - RB + (RP FN) - JE
\]

The interpretation of the terms in parentheses will be given later.

For the AC design:

\[
SB(AC) = -(SFpTN + SFpTN) - FP TN - (RFpTN + R NFpTN)
\]

In terms of the earlier appendix definitions,

\[
SB(AC) = SB - (SFpTN) - RB - (RFpTN) + JE
\]

The parenthetical terms in eqs. (1) and (2) reflect the interpretation problems unique to the one-directional designs. These are termed design effects on misclassifications and bear a direct relationship to random error in the observations.

Survey random error is defined to be the rate of compensating survey errors:

\[
SRE = SFpTN + SFpTN - SFpTN - SFpTN
\]

Record random error is defined similarly:

\[
RRE = RFpTN + RFpTN - RFpTN - RFpTN
\]

**EXHIBIT 7 -- EXPRESSIONS FOR SURVEY AND RECORD RANDOM ERROR BY TRUE ERROR CHARACTERISTIC**

<table>
<thead>
<tr>
<th>True Error Characteristic</th>
<th>SRE = Survey Random Error</th>
<th>RRE = Record Random Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1--Only Random Error (FP rate = FN rate&gt;0)</td>
<td>( 2 SFp PN ) or ( 2 SFp FN )</td>
<td>( 2 FP p PN ) or ( 2 FP p FN )</td>
</tr>
<tr>
<td>Case 2--Net Positive Bias (FP rate &gt; FN rate)</td>
<td>( 2 SFp TN )</td>
<td>( 2 SFp FN )</td>
</tr>
<tr>
<td>Case 3--Net Negative Bias (FN rate &gt; FP rate)</td>
<td>( 2 SFp TN )</td>
<td>( 2 SFp FN )</td>
</tr>
</tbody>
</table>
The expressions for random error may be rewritten as in Exhibit 7, depending on the relative rates of false positives and negatives.

The design effect on survey misclassification for the AB design is $S_{FN}T_p$ (eq. 1). For cases 1 and 2, reference to Exhibit 7 shows that this is equal to $1/2$ SRE. For case 3, it may be written as $(S_{FP}T_F - S_{FP}T_N) + S_{FP}T_N = -SB + 1/2$ SRE.

The design effect on record misclassifications for the AB design is $R_{FP}T_N$ (eq. 1). For cases 1 and 2, the effect is equal to $1/2$ RRE. For case 3, it may be expressed as $(R_{FP}T_F - R_{FP}T_N) + R_{FP}T_N = +RB + 1/2$ RRE.

The interpretation of the design effects is that the AB design includes half the survey random error and half the record random error in its definition of survey bias. Furthermore, if the survey contains a net negative bias and/or the records contain a net positive bias, the AB design will miss them completely.

The design effect on survey misclassifications for the AC design is $-S_{FP}T_N$ (eq. 2). For cases 1 and 3, it equals $-1/2$ SRE, causing the estimate of survey bias to be more negative. For case 2, the effect is written as $-SB - 1/2$ SRE, indicating that a true positive net survey bias will be undetected and a negative bias equal to half the survey random error rate will be substituted. The design effect on record misclassifications for the AC design is $-R_{FP}T_p$ (from eq. 2). The AC design includes $1/2$ of the record random errors in the definition of survey bias and cannot detect a true net negative record bias.

A summary of the error elements in the survey bias definitions of each of the three record check designs is in Exhibit 8.

**EXHIBIT 8 -- ELEMENTS OF ERROR INCLUDED IN RECORD CHECK DEFINITIONS OF SURVEY BIAS**

<table>
<thead>
<tr>
<th>Type of Record Check Design</th>
<th>True Survey Bias</th>
<th>Error Elements Included in Survey Bias Definitions</th>
<th>Design Effect on Survey Misclassification</th>
<th>Record Bias Effect</th>
<th>Design Effect on Record Misclassification</th>
<th>Joint Error Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>+</td>
<td>None</td>
<td>-</td>
<td>-</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>AB</td>
<td>+</td>
<td>Cases 1 and 2: +1/2 SRE</td>
<td>-</td>
<td>Cases 1 and 3: +1/2 RRE</td>
<td>Case 2: +RB + 1/2 RRE</td>
<td>-</td>
</tr>
<tr>
<td>AC</td>
<td>+</td>
<td>Cases 1 and 3: -1/2 SRE</td>
<td>-</td>
<td>Cases 1 and 2: -1/2 RRE</td>
<td>Case 3: +RB - 1/2 RRE</td>
<td>+</td>
</tr>
</tbody>
</table>