Validation of Outcomes Through the Analysis of the Process-Outcome Relation: Limitations

LETICIA KRAUSS SILVA

Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Ministry of Health, Brazil

Validation of outcomes of different levels of neonatal care through the analysis of the relationship between process and outcome of care was attempted as part of an effectiveness study and quality assurance programme. Adequacy rates and outcome measures were calculated for second- and third-level units and centres according to the babies' level of risk. Differences in adequacy rates were apparently influenced by the differential distribution of babies by care level sectors among maternity centres. The comparison of mortality rates for "treated" and "untreated" subgroups did not show higher mortality rates for the "untreated" subgroups. This finding could be related to the phenomenon of "confounding by indication". Consistency of differences in adequacy rates with differences in general outcomes was observed for most subgroups analysed; lack of consistency could be related to low effectiveness. Indications of low effectiveness related to secondary and more so to tertiary level technologies were observed. The study approach was found more useful for comparisons made at the level of maternity services as a whole, where "confounding by indication" does not operate than at the level of sectors.

Key words: Validation of outcomes, process-outcome relation, effectiveness analysis.

INTRODUCTION

Analysis of the process of hospital care frequently means dealing with a complex set of medical technologies. Analysis of outcomes is further burdened by the scarcity of efficacy figures provided by well-designed studies, by the low efficacy and specificity (in relation to outcomes) of many procedures and by the presence of confounding factors, leading to difficulties in establishing the "attributional validity" of individual/group procedures [1,2]. In order to obviate such obstacles, the analysis of the relation between process and outcome has been suggested to minimize the uncertainties around the "attributional validity" associated with the procedures under study [1]. Adequate approaches for producing valid estimates of effectiveness of medical care, using non-experimental data, are presently an area of considerable debate and research [3-9].

Several sources of confounding are expected when performing analysis of medical care and, in relation to non-experimental data, techniques to control for different sources of bias are far from perfect [10-15]. In relation to "confounding by indication" for treatment, according to Greenland and Neutra [16], it is a natural consideration in the assessment of medical technology because it is usually guaranteed in non-experimental studies. The reason given is that "indications for treatment are almost always indicators of special risk and, therefore, persons who are selected for special treatment will automatically be at different risks from those who receive the more standard treatment" (p.361). Miettinen [17], agreeing with the relevance of the problem, considers that the research into the efficacy, or "intended effects", of interventions is an outstanding area for the occurrence of confounding, particularly of "confounding by indication". He explains: "Interventions are commonly prompted by an indication, a state or event that signifies the prospect of an untoward outcome. Thus, by the very rationality of decisions to intervene, the treated tend to differ from the untreated with respect to their outlooks for the outcome criterion in efficacy assessment; there tends to be "confounding by the indication" - usually such that the treated tend to have less favourable outcome than the untreated" (p. 13). However, there is controversy about ways to deal with the problem. A careful control by analysis was particularly recommended by Greenland and Neutra in relation to "confounding by indication" in evaluating the benefit of a...
importance of controlling such bias in the analysis. They conclude by emphasizing the need for the control of diagnostic technology.\(^2\) Poses et al. [18], evaluating a therapeutic technology, reinforce the need for the control by analysis of prognostic variables, particularly of those associated with case severity. On the other hand, for Miettinen, such confounding "is often difficult or impossible to deal with in nonexperimental terms".

First reason: "the characterization of initial indication is often quite nebulous, even if done with great care". Second reason: when continuation of the treatment (or the initially untreated remaining untreated) is dependent on the apparent response, "this feedback creates a select set of patients". In relation to such select patients, the author comments: "Their intrinsic prospects with respect to the outcome criteria are not implicit in the initial indication for intervention and the subsequent indications for continuation tend not to be recorded in a comparable fashion between the compared groups" (p.44). His conclusion about the problem is that "non-experimental research is generally better suited for the study of unintended effects than the efficacy of particular interventions modalities, practices, or programs" (p.44).

The present study was part of a broader project designed to determine the cost-effectiveness of different levels of neonatal care in the city of Rio de Janeiro in relation to babies of different levels of risk/need [8] and to produce information for a quality assurance programme. It intended to explore approaches for the analysis of the relation process outcome as a way to validate outcomes (effectiveness) of care.

**METHODOLOGY**

An information system was developed to produce data on risk/need, process and outcomes related to different levels of neonatal care provided at the public maternity services in Rio de Janeiro. The selection of procedures to be included in the neonatal summary was based on scientific evidence of efficacy in relation to relevant neonatal problems, in terms of mortality rates. Process-outcome pairs (selected procedure—mortality rate related to its indications) with established "causal validity" were specified as the basis to carry out the analysis of effectiveness [19] (Table 1). The present study refers to a population of approximately 36 000 babies without major congenital anomalies, born at maternity services of secondary, intermediate and tertiary levels of care, who were cared for at different care level sectors—considering the highest care level as the reference—from January 1987 to June 1989 (Table 2).\(^3\) Only later neonatal care—meaning the care delivered after labour room care, in neonatal or rooming in facilities, up to discharge—was considered by the study. Babies who died in the labour room were therefore excluded from the study. Babies were classified by levels of risk—low, intermediate, high, very high.

---

\(^2\)The authors exemplify with lower and with higher risk patients being selected for the performance of particular new procedures, in both cases, a bias being produced in the estimation of effect on a simple comparison of the treated and the untreated groups. They conclude by emphasizing the importance of controlling such bias in the analysis.

---

**TABLE 1. Selected process–outcome pairs with established “causal validity” used in the study.**

<table>
<thead>
<tr>
<th>Referent</th>
<th>Procedure*</th>
<th>Outcome</th>
<th>Scientific evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small for Gestational Age</td>
<td>Glycemia determination within 4 hours of birth (and correction of hypoglycemia)</td>
<td>Decreased neurological morbidity and mortality</td>
<td>Neligan et al., 1963 [30]</td>
</tr>
<tr>
<td>(SGA)</td>
<td></td>
<td></td>
<td>Brown et al., 1963 [31]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Shelley, 1964 [32]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rusel and McKay, 1966 [33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Usher, 1962 [34]</td>
</tr>
<tr>
<td>Respiratory Distress Syndrome</td>
<td>Venous hydration within 6 hours of birth (physiologic saline with glucose and sodium bicarbonate)</td>
<td>Decreased mortality, especially among babies with birthweight less than 1500g</td>
<td>Auld et al., 1966 [35]</td>
</tr>
<tr>
<td>(RDS)</td>
<td></td>
<td></td>
<td>Cornblath et al., 1966 [36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kitchen et al., 1971 [37]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hobel et al., 1972 [38]</td>
</tr>
<tr>
<td>Respiratory disorders</td>
<td>Chest X-rays within 6 hours of birth</td>
<td>Improved diagnostic precision (and improved prognosis)</td>
<td>Bauman and Nadelhaft, 1958</td>
</tr>
<tr>
<td>(transient tachypnea,</td>
<td></td>
<td></td>
<td>Weintraub et al., 1966 [40]</td>
</tr>
<tr>
<td>respiratory distress syndrome,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pneumonia, pneumothorax)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory disorders</td>
<td>Mechanical ventilation/continuous positive airway pressure</td>
<td>Decreased mortality, especially for those babies with low birthweight/gestational age</td>
<td>Rhodes et al., 1973 [41]</td>
</tr>
<tr>
<td>(severe respiratory distress</td>
<td></td>
<td></td>
<td>Allen et al., 1977 [42]</td>
</tr>
<tr>
<td>syndrome, pneumonia, apnea</td>
<td></td>
<td></td>
<td>Murdoch et al., 1977 [43]</td>
</tr>
<tr>
<td>repetens)</td>
<td></td>
<td></td>
<td>Manginello et al., 1978 [44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lindroth et al., 1980 [45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Greenough and Robertson, 1983 [46]</td>
</tr>
</tbody>
</table>

*The process of care for those babies included other components whose influence was controlled by the studies.
TABLE 2. Total numbers and percentages of inpatient births, non-transferred excluding major congenital anomalies and labour room deaths

<table>
<thead>
<tr>
<th></th>
<th>ML III</th>
<th>ML Int</th>
<th>ML II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>7427(81.9)</td>
<td>4786(79.0)</td>
<td>17238(83.7)</td>
</tr>
<tr>
<td>Med</td>
<td>1119(12.3)</td>
<td>986(16.3)</td>
<td>1987(9.7)</td>
</tr>
<tr>
<td>High</td>
<td>193(2.1)</td>
<td>133(2.2)</td>
<td>269(1.3)</td>
</tr>
<tr>
<td>Ext</td>
<td>77(0.8)</td>
<td>60(1.0)</td>
<td>174(0.8)</td>
</tr>
<tr>
<td>Ign</td>
<td>257(2.8)</td>
<td>90(1.5)</td>
<td>916(4.5)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9073</td>
<td>6055</td>
<td>20584</td>
</tr>
</tbody>
</table>

The population from which the sample was drawn includes inpatient births, non-transferred, excluding major congenital anomalies and labour room deaths ($n=35,712$ babies).

ML III—maternity services delivering up to secondary care level.

ML Int—maternity services delivering up to intermediate care level.

ML II—maternity services delivering up to secondary care level.

medium, high and extreme—according to an algorithm consisting basically of the variables birthweight, gestational age and 5th minute Apgar score, which were selected through a prior analysis of confounding of 10 risk factors listed in the literature in relation to overall maternity services outcomes (taking into account all care level sectors). A small fraction of babies could not be classified by risk level due to lack of relevant information; more than 80% of those babies were cared for at the rooming-in sector in all maternity services subgroups. They were excluded from the study. Low-risk babies were also excluded from this study, as selected procedures are not generally indicated nor are the adverse outcomes analysed observed in this group. Extensive validation procedures were performed in relation to data on risk level, care level and outcomes. Mortality rates were corrected for the transfer of newborn babies, using the mortality rates observed for babies whose transfer was indicated but not carried out in order to make mortality rates of different maternity service care levels comparable. Figure 1 shows how the elements used in this study are related and the way they were specified.

The study focused on approaches to validating outcomes as “related” to procedures provided to patients by way of analysing the relation between use of a procedure when it was indicated (adequacy) and supposedly “related outcomes”. The specification of the referent—a diagnosis or condition to which standards of care apply [1]—was based on the criteria for including babies by the efficacy studies of the related procedures.

Description of the comparison

An analysis of outcomes preceded the analysis of the relation between process and outcomes which is the object of the present study.

3 Neonatal technologies were classified [20] into labour room assistance (early care) and three levels of care (later care) which constitute a hierarchy by successive inclusions: Labour-room assistance: including procedures for newborn baby resuscitation, observation, first detailed examination plus screening. Primary Care Level Sector—Rooming in, including relatively simple procedures, e.g., phototherapy, antibiotic therapy, eventual laboratory exams (hematocrit, hemogram, glycemia, bilirubinemia), and X-rays as well as screening for syphilis, phenylketonuria and hypothyroidism. Secondary Care Level Sector—Temperature control (incubator), oxygen therapy (with oxihood and ambient oxygen analyser), venous hydration, radiography, laboratory control of glycemia, hemogram/hematocrit and bilirubinemia, electrolytes, heart rate and apnea monitoring, nasogastric tube feeding, arterial gases monitoring, liquor examination, blood transfusion and screening tests. Intermediate Care Level Sector—Assisted ventilation (continuous positive airway pressure and intermittent mandatory ventilation techniques) and continuous monitoring/nursing supervision by skilled nursing and medical staff. It also incorporates secondary level technologies. Tertiary Care Level Sector—Assisted ventilation (same techniques used at the intermediate sector but for longer term), total parenteral nutrition and multiple and continuous monitoring/nursing supervision by skilled nursing and medical staff. Post-surgical care of major surgical procedures (and its corresponding referent) had to be excluded from the analysis due to missing records. The sector also comprehends secondary level technologies. It is important in this context to make distinctions between care level sector and care level of maternity service. The first means a defined set of procedures which usually take place in a specific physical unit or sector. Care level of maternity service encompasses the service or centre as a whole and corresponds to the highest care level it delivers [21,22], as maternity services usually include more than one care level sector.

4 The risk levels are described below [20]: Low risk: newborn babies with the following characteristics: birthweight higher than 1999 g, 5th min. Apgar score higher than the 6 and ICD codes compatible with rooming-in care; birthweight higher than 1899, gestational age equal or higher than 34 weeks, 5th min. Apgar score higher than 6 and ICD codes compatible with rooming-in care. Medium risk: birthweight from 1500 to 1899 g and gestational age equal to or higher than 32 weeks and 5th min. Apgar score higher than 3; birthweight from 1900 and 1999 g, gestational age less than 34 weeks and 5th min. Apgar score higher than 3; birthweight equal to or more than 2000g disregarding the gestational age, 5th min. Apgar score equal to 2 or 3 (more than 999g and more than 29 weeks of gestation). ICD code 768.5 (severe perinatal asphyxia), when 5th min Apgar score is ignored. Extreme risk: birthweight less than 1000 g or gestational age less than 29 weeks and 1 day (somatic Capurro’s lower limit) or 5th min Apgar score lower than 2 or ICD code 765.0 (extreme prematurity), when gestational age is ignored and birth weight less than 1000g or Rh and non-Rh related hydrops (at birth) (ICD codes 773.3 and 778.0).

5 Mortality rates of a referring centre were corrected by assuming an “expected” rate of mortality for the transferred out babies as if they had not been transferred. To do that, transferred out babies, after exclusion of cases with major congenital anomalies, were stratified by risk level, each stratum assuming the mortality observed for the corresponding class of risk of those newborn babies whose transfer was indicated but not realized (for unavailability of intensive care beds).
Outcomes

After considering the findings of the above-mentioned analysis of confounding in regard to maternity services outcomes, and effecting the steps to avoid other sources of bias, later neonatal mortality rates (in-hospital mortality excluding labour room deaths) were then calculated for each risk level by maternity service and sector care level subgroups. Observed differences were checked for statistical significance by using a test for comparison of proportions [23]. The confidence level used was 95%, observing the two-sided value. The secondary care level centres (maternity services) subgroup was considered as the reference category for the comparison of maternity services, the intermediate care sector subgroup was the reference category for the comparison between tertiary care sector and intermediate care sector rates [24]. The study did not elaborate/utilize standards of outcomes (e.g. expected mortality rates) for individual or sets of technologies in evaluating services.

Comparison of adequacy rates. The rate of performance of selected procedures in the presence of related conditions/risk factors was computed so as to determine the percentage of patients who received care that they needed (e.g. the percentage of babies with selected respiratory conditions who were submitted for chest X-rays within four hours of birth).

To calculate the adequacy rate of a group of procedures, as in the group of secondary sector procedures, a simple mean of the individual rates was utilized. Observed differences among different maternity service and sector levels subgroups on grouped procedures, by risk level, were checked for statistical significance by using a test for comparison of proportions [23].

Analyses of the relation between process and outcome

Comparison of mortality rates for “treated” and “untreated” groups: cases were divided into “treated” and “untreated” categories, depending on whether or not they had received the selected indicated procedures used to signify adequacy. The results in the two categories were compared as judged by the occurrence of deaths. The objective of the analysis was to verify the hypothesis that mortality rates would be greater for the “untreated” groups of babies, as compared to the “treated” groups.

Comparison of relative risks of not receiving appropriate care (selected procedures) when indicated: The objective of this analysis was to verify the hypothesis that relative risks, in terms of outcomes, of not receiving a “treatment” would be approximately equal for different maternity service and sector levels by risk level. Observed differences among maternity services and sectors were checked for statistical significance using a test derived from the two-binomial model [24].

Analysis of the consistency of differences in average rates of adequacy with differences in general outcomes: the extent to which differences in average rates of adequacy were consistent with corresponding differences in total mortality rates between maternity service care levels and between sector care level subgroups, according to risk level, was analysed—consistency meaning a
and intermediate maternity services for extreme-risk
The difference in mortality rates between tertiary services
for a similar risk level and maternity service subgroup.
mortality rates than tertiary and intermediate sectors
for the latter. Secondary sectors presented lower
secondary services resulted in highly unfavourable rates
tertiary and intermediate maternity services against
secondary level sector mortality rates between both
being highly significant statistically, except for extreme-
procedure and average adequacy for grouped procedures
(secondary and tertiary procedures, that is, mechanical
ventilation and total parenteral nutrition) by risk level,
procedure and average adequacy for grouped procedures
subgroups, should be consistent with the corresponding
secondary sector mortality rates, in the sense that more
favourable rates of adequacy should be observed in the
centres which presented the best results with respect to
mortality.
The study considered two general alternative assumptions: (a) patients within each risk level were homo-
geneous, taking a sector or a service as the unit for
analysis and no selection occurred, (b) patients within
each risk level were not homogeneous and indication for
treatment produced a selection bias. The implications of
both assumptions for the proposed analyses are schemati-
cally described in Appendices 1 and 2.

RESULTS
Outcomes
Tables 3 and 4 show outcomes—corrected and
uncorrected figures—calculated according to risk level
by maternity service and sector subgroups. They were
used as a reference for the analysis of the relation between
process and outcome.
Maternity services mortality (Table 3). Tertiary centres
presented favourable results for later care—in terms of
overall rates, that is, of rates related to all babies of a
similar risk level, including all sector levels—when
compared to secondary maternity services, differences
being highly significant statistically, except for extreme-
risk babies (p = 0.10). Intermediate care maternity
services presented lower rates in relation to secondary
maternity services for high-risk babies (p = 0.06). In
relation to intermediate maternity services, tertiary
services presented statistically superior overall rates
only for medium-risk babies.
Sector mortality (Table 4). The comparison of
secondary level sector mortality rates between both
tertiary and intermediate maternity services against
secondary services resulted in highly unfavourable rates
for the latter. Secondary sectors presented lower
mortality rates than tertiary and intermediate sectors
for a similar risk level and maternity service subgroup.
The difference in mortality rates between tertiary services
and intermediate maternity services for extreme-risk
babies cared for at secondary sectors was also striking.

| TABLE 3. Later neonatal mortality (%) by risk level and
| maternity service care level: corrected and uncorrected rates* |
| Level of | ML III (all sectors) | ML Int (all sectors) | ML II (all sectors) |
| risk     |                  |                  |                  |
| Medium   | 3.0f (2.9)       | 5.0 (4.0)        | 5.6 (4.3)        |
| High     | 29.3f (28.5)     | 33.1§ (31.6)     | 42.7 (37.2)      |
| Extreme  | 85.2f (84.4)     | 92.1 (91.7)      | 91.9 (90.8)      |

*The sample includes inpatient births, non-transferred,
excluding major congenital anomalies and labour room
deaths. Correction was effected to take into account expected
mortality related to the transferred out babies. Uncorrected
figures appear in parentheses. The sample size, including
transferred out babies, was 5153 (corrected figures); the
exclusion of transferred out babies reduced the sample to 4998
babies (uncorrected figures).
For explanation of table headings see Table 2.
Comparison with ML II: t p = 0.001 t p = 0.003 § p = 0.06
||p = 0.10.
†Comparison with ML Int: p = 0.02.

| TABLE 4. Later mortality by care level sector according to
| risk level and maternity service care level: corrected and
| uncorrected figures* |
| Level of | Secondary sector | Tertiary sector |
| risk     | ML III | ML Int | ML II | ML III | ML Int |
| Medium   | 1.8†  | 3.4|| | 8.5 | 31.8f (31.8) | 57.0 (50.0) |
| High     | 9.8†  | 14.1§ | 42.7 | 57.4** (56.4) | 74.4 (72.5) |
| Extreme  | 50.6§ | 92.9 | 91.9 | 95.2 (94.9) | 91.4 (90.6) |

*The sample includes inpatient births, non-transferred,
excluding major congenital anomalies and labour room
deaths. Tertiary and intermediate sector rates were corrected
for transferred out babies. Uncorrected figures appear in
parentheses. The sample size, including transferred out babies,
was 3727 (corrected figures); the exclusion of transferred out
babies reduced the sample to 3572 babies (uncorrected figures).
For explanation of table headings see Table 2.
†Comparison with ML II: p < 0.001, and with ML Int:
p = 0.07.
‡Comparison with ML II: p < 0.001.
§Comparison with ML II and ML Int: p < 0.001.
Comparison with ML II: || p < 0.01
Comparison with ML Int: §p = 0.006 **p = 0.06.

Analysis of process
An overall picture of the actual utilization of the
facilities under study (later care) is shown in Table 5. It
lays out the percentages of babies by risk level (including
low-risk babies), sector level and maternity service level
subgroup. It shows sizeable and different proportions of
high- and extreme-risk babies who were treated at
secondary level sectors in tertiary and intermediate
maternity services.

Adequacy analysis. Adequacy rates for each selected
procedure and average adequacy for grouped procedures
(secondary and tertiary procedures, that is, mechanical
ventilation and total parenteral nutrition) by risk level,
Service level was the same for both procedures as they were alternative treatment procedures for a same referent.

It must be observed that the mortality rate related to the "untreated" group for hood oxygen or mechanical ventilation in the analysis by maternity subgroups. The findings were inverse to the hypothesis, not hold in general for maternity service and sector level.

The hypotheses that adequate rates were consistent with the occurrence of a bias in the indication of patients for treatment, or both. The fact that for both medium and high-risk babies the hypothesis was not supported could admit two explanations according to Appendices 1 and 2: (a) a low effectiveness on the part of the service, (b) a bias in the indication of patients for treatment, or both. The fact that for both medium and high-risk babies the hypothesis was not supported could admit two explanations according to Appendices 1 and 2.

Analysis of the relation between process and outcome

Comparison of mortality rates for "treated" and "untreated" groups

Mortality rates (%) for "treated" and "untreated" subgroups by maternity service and sector levels for all risk levels are shown in Table 8 for high-risk babies at tertiary maternity services.

The analysis related to the hypothesis—"untreated" groups would have a higher mortality rate than the "treated" groups—was limited by the occurrence of two problems: mortality rates related to small numbers and absent mortality rates due to lack of subjects (generally in relation to the "untreated" group, except for total parenteral nutrition). Despite those problems, the inspection of rates made it apparent that such a hypothesis did not hold in general for maternity service and sector level subgroups. The findings were inverse to the hypothesis, for medium- and high-risk babies, in the analysis of maternity services subgroups (considering all sectors). It was generally opposed by medium-risk subgroups and generally held true for extreme-risk subgroups. Also, for nearly 25% of the subgroups, the findings were contradictory in relation to the hypothesis. There was no tendency for supporting or refusing the hypothesis in association with level of procedure (secondary or tertiary) or to maternity service level.

The fact that the hypothesis was not generally supported could admit two explanations according to Appendices 1 and 2: (a) low effectiveness on the part of the service, (b) a bias in the indication of patients for treatment, or both. The fact that for both medium and high-risk babies the hypothesis was not supported could admit two explanations according to Appendices 1 and 2.

Comparison of relative risks of not receiving care when indicated: With respect to the hypothesis that relative risks of not receiving "treatment" were equal, a further problem emerged: a significant proportion of rates being equal to zero. All above-mentioned problems with rates disturbed this analysis, as comparison of risk ratios among maternity services groups and care level sectors for individual procedures depended upon valid rates for each of four to six mortality rates, which were frequently unavailable. Nevertheless, whenever possible, relative risks were calculated. Comparison of relative risks showed that they were not equal. No particular trend of supporting or opposing such hypothesis was observed either for maternity service or sector level.

These results indicate a bias in the selection of patients, if performance of procedures by study services were satisfactory, according to Appendices 1 and 2.

---

TABLE 5. Distribution of babies by sector care level (%) according to risk level and maternity service care level

<table>
<thead>
<tr>
<th>Level of risk</th>
<th>Primary sector</th>
<th>Secondary sector</th>
<th>Tertiary/Int. sector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML III</td>
<td>ML Int</td>
<td>ML II</td>
</tr>
<tr>
<td>Low</td>
<td>92.6</td>
<td>94.5</td>
<td>95.2</td>
</tr>
<tr>
<td>Medium</td>
<td>34.7</td>
<td>33.4</td>
<td>35.7</td>
</tr>
<tr>
<td>High</td>
<td>1.6</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Extreme</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*The population from whom the sample was drawn includes inpatient births, non-transferred, excluding major congenital anomalies and labour room deaths (n = 34,449 babies). Babies were classified according to the highest level of care they received during hospitalization.

Percentages are related to the row totals for each maternity service level subgroup.

For explanation of table headings see Table 2.

Comparison with ML II: †p < 0.001 ‡p = 0.007.
Comparison with ML Int: §p < 0.001 †p = 0.09 ‡p = 0.04 **p = 0.007.
Comparison with ML Int: ††p = 0.02 ‡‡p = 0.06 §§p = 0.001.

According to maternity service and sector levels with corresponding p values, are shown in Tables 6 and 7, for high-risk babies only.

Overall adequacy rates were statistically higher for secondary level technologies at tertiary maternity services than at secondary level maternity services and at intermediate maternity services, for all risk levels. Adequacy rates for tertiary level technologies were higher at tertiary maternity services than at intermediate maternity services for all risk levels. Corresponding differences between intermediate and secondary level maternity services were significant only for extreme-risk babies. Statistically significant differences in average adequacy rates were almost always consistent with differences in individual procedures, by maternity service and by sector level for each risk level considered.

It must be observed that the mortality rate related to the "untreated" group for hood oxygen or mechanical ventilation in the analysis by maternity service level was the same for both procedures as they were alternative treatment procedures for a same referent.
TABLE 6. Adequacy rates by maternity service care level (%). Later care, high-risk babies

<table>
<thead>
<tr>
<th>Conditions vs technologies</th>
<th>Tertiary</th>
<th>Intermed.</th>
<th>Secondary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small for gestational age vs glycemia</td>
<td>90.4</td>
<td>89.5</td>
<td>67.5</td>
</tr>
<tr>
<td>Apgar score vs glycemia</td>
<td>88.0</td>
<td>71.2</td>
<td>55.5</td>
</tr>
<tr>
<td>Respiratory Distress Syndrome/low Apgar score vs venous hydration</td>
<td>89.7</td>
<td>81.2</td>
<td>83.1</td>
</tr>
<tr>
<td>Respiratory conditions vs chest X-rays</td>
<td>86.3</td>
<td>59.6</td>
<td>50.4</td>
</tr>
<tr>
<td>Prolonged ruptured membranes vs hemogram</td>
<td>90.7</td>
<td>50.0</td>
<td>73.7</td>
</tr>
<tr>
<td>Hematological disorders/sepsis vs exchange transfusion</td>
<td>33.3</td>
<td>21.0</td>
<td>12.1</td>
</tr>
<tr>
<td>Respiratory disorders vs hood oxygen therapy</td>
<td>34.3</td>
<td>31.6</td>
<td>67.1</td>
</tr>
<tr>
<td>Respiratory disorders vs mechanical ventilation (including continuous positive airway pressure)</td>
<td>53.3</td>
<td>41.8</td>
<td></td>
</tr>
<tr>
<td>Sepsis/necrotising enterocolitis/mechanical ventilation vs total parenteral nutrition</td>
<td>12.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Average (%):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary technologies</td>
<td>79.7†</td>
<td>62.1</td>
<td>57.1</td>
</tr>
<tr>
<td>(excl. hood oxygen therapy)</td>
<td>†</td>
<td>†</td>
<td>†</td>
</tr>
<tr>
<td>Tertiary technologies*</td>
<td>32.7§</td>
<td>20.9</td>
<td>–</td>
</tr>
<tr>
<td>All technologies (excluding total parenteral nutrition)</td>
<td>70.7¶</td>
<td>55.8</td>
<td>51.2</td>
</tr>
</tbody>
</table>

The sample includes inpatient births, non-transferred, excluding major congenital anomalies and labour room deaths (n = 595 babies).

* Tertiary technologies—mechanical ventilation and total parenteral nutrition
†/‡ Comparison with secondary and intermediate services: p < 0.001.
§ Comparison with intermediate services: p < 0.001.
¶ Comparison with secondary services: p < 0.001.

Analysis of the consistency of differences in average rates of adequacy with differences in general outcomes

For the overwhelming majority of subgroups analysed (maternity service and sector levels by risk level, n = 27), consistency between differences in average rate of adequacy (Tables 6 and 7) and in total mortality rates (Tables 3 and 4) was observed. Differences were statistically significant at a p level less than 0.05 and less than or equal to 0.10 for both mortality and corresponding adequacy rates—for nearly 45% and 60% of the subgroups, respectively. Consistency was low—meaning that, though rate differences were consistent, either adequacy or mortality rate, rarely both, presented p values higher than 0.10—for about 25% of the subgroups. For two subgroups, consistency was not observed as the statistically significant difference observed in one kind of rate was paired by a near zero difference in the other kind of rate. Also, for two other subgroups, differences in rate differences were inconsistent: only the mortality rate difference was statistically significant, while the difference in adequacy rates was rather small. These cases occurred in the comparison of secondary level sectors between intermediate and secondary level maternity services, for medium- and high-risk babies.

Differences related to medium-risk babies tended to present high consistency as well as differences between tertiary and secondary level maternity services. Secondary level procedures showed somewhat more consistency than tertiary procedures (mechanical ventilation and total parenteral nutrition). Differences between intermediate maternity services and secondary level maternity services tended to present lower consistency, including also the above-mentioned case of no consistency.

The observed tendency towards consistency was expected for both assumptions related to homogeneity of babies regarding severity within the same risk level (Appendices 1 and 2).

DISCUSSION

The present study explored ways of analysing the relation between outcomes and process of care—measured through adequacy rates of procedures—for differ-
TABLE 7. Adequacy rates by sector care level according to maternity service level (%). Later care, high risk babies

<table>
<thead>
<tr>
<th>Conditions vs technologies</th>
<th>Secondary sector</th>
<th>Tertiary sector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML III</td>
<td>ML Int</td>
</tr>
<tr>
<td>Small for gestational age vs glycemia</td>
<td>91.9</td>
<td>86.7</td>
</tr>
<tr>
<td>Low Apgar score vs glycemia</td>
<td>82.4</td>
<td>56.7</td>
</tr>
<tr>
<td>Respiratory distress syndrome/low Apgar score vs venous hydration</td>
<td>86.8</td>
<td>71.1</td>
</tr>
<tr>
<td>Respiratory conditions vs chest X-rays</td>
<td>83.3</td>
<td>48.3</td>
</tr>
<tr>
<td>Prolonged ruptured membranes vs hemogram</td>
<td>81.0</td>
<td>40.9</td>
</tr>
<tr>
<td>Hematological disorders/sepsis vs exchange transfusion</td>
<td>20.0</td>
<td>10.8</td>
</tr>
<tr>
<td>Respiratory disorders vs hood oxygen therapy</td>
<td>61.8</td>
<td>56.9</td>
</tr>
<tr>
<td>Respiratory disorders vs mechanical ventilation (including continuous positive airway pressure)</td>
<td>100.0</td>
<td>25.6</td>
</tr>
<tr>
<td>Sepsis/necrotising enterocolitis/mechanical ventilation vs total parenteral nutrition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average (%): secondary technologies 72.5†

Average (%): Secondary technologies (excluding hood oxygen therapy) 84.0 77.6

Tertiary technologies 62.8 50.0

The sample includes inpatient births, non-transferred, excluding major congenital anomalies and labour room deaths (n = 595 babies).

For explanation of table headings see Table 2.

*Comparison with ML II: p = 0.001.
†Comparison with ML Int: p = 0.004.

ent levels of neonatal care, with the objective of validating observed outcomes. Procedures were selected on the basis of scientific evidence of efficacy for relevant problems in terms of neonatal mortality. The analyses were restricted by the limitations (in terms of precision) of available data on discharge diagnoses, by the limitation in the scope of procedures recorded, by the scarcity of studies on the efficacy and effectiveness of neonatal technologies to be used as standards of performance as well as by the low sensitivity/specificity estimated for the procedures studied [25–29]. Special attention was devoted to the possible occurrence of confounding by indication, in the analysis of the process–outcome relation.

Classifying patients by severity—though crucial for the analysis of outcomes—is not a simple matter [47], even considering that the related discharge summary was designed for analysing effectiveness. Delimiting referents, including severity status, though fundamental for effectiveness analysis, should also be considered a difficult endeavour [1], especially if one is trying to obtain uniform and valid information on thousands of babies from different services [48]. Nevertheless, the present study classification of risk should be considered sound as it was preceded by an analysis of confounding of 10 relevant risk factors and was based on well-known powerful prognostic variables. Furthermore, differences in total mortality rates among risk level subgroups were high at all maternity service levels (Tables 3 and 4). However, referents for each studied procedure were probably not sufficiently specified—though they comprehended the specific disease/condition related to each procedure and more general risk variables associated with risk level—and consequent lack of homogeneity of patients probably occurred within a similar sector and among sectors of a similar maternity service subgroup, giving room to "confounding by indication". The existence of babies classified as high risk and as extreme risk who could survive under a "low technology" regimen (Table 4) indicates some degree of misclassification of babies by the study.

Correcting outcomes for differences in rates of transfer was a necessary step for making the study services comparable. The method used by the present study for effecting such correction resulted in a substantial difference in mortality rates for the secondary maternity services subgroup, as expected, except for extreme-risk babies and also for the intermediate maternity services,
TABLE 8. Mortality (%) related to treated and untreated patients at tertiary maternity services by sector care level. Later care, high-risk babies

<table>
<thead>
<tr>
<th>Conditions vs technologies</th>
<th>Secondary plus tertiary sectors</th>
<th>Secondary sector</th>
<th>Tertiary sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small for gestational age vs glycemia</td>
<td>17.0 (47)</td>
<td>2.9 (34)</td>
<td>53.8 (13)</td>
</tr>
<tr>
<td>Low Apgar score vs glycemia</td>
<td>37.9 (65)</td>
<td>7.1 (28)</td>
<td>62.2 (37)</td>
</tr>
<tr>
<td>Respiratory distress syndrome/low Apgar score vs venous hydration</td>
<td>40.2 (87)</td>
<td>9.1 (33)</td>
<td>59.3 (54)</td>
</tr>
<tr>
<td>Respiratory conditions vs chest X-rays</td>
<td>36.2 (69)</td>
<td>10.0 (30)</td>
<td>56.4 (39)</td>
</tr>
<tr>
<td>Prolonged ruptured membranes vs hemogram</td>
<td>41.0 (39)</td>
<td>17.6 (17)</td>
<td>59.1 (22)</td>
</tr>
<tr>
<td>Hematological disorders/sepsis vs exchange transfusion</td>
<td>63.6 (11)</td>
<td>33.3 (3)</td>
<td>75.0 (8)</td>
</tr>
<tr>
<td>Respiratory disorders vs hood oxygen therapy</td>
<td>5.9 (35)</td>
<td>5.9 (35)</td>
<td>58.9 (56)</td>
</tr>
<tr>
<td>Respiratory disorders vs mechanical ventilation</td>
<td>58.9 (56)</td>
<td>58.9 (56)</td>
<td>-</td>
</tr>
<tr>
<td>Sepsis/necrotising enterocolitis vs Mechanical ventilation</td>
<td>35.0 (20)</td>
<td>31.7 (145)</td>
<td>35.0 (20)</td>
</tr>
<tr>
<td>Total parenteral nutrition</td>
<td>-</td>
<td>-</td>
<td>63.8 (58)</td>
</tr>
</tbody>
</table>

The sample includes inpatient births, non-transferred, excluding major congenital anomalies and labour room deaths (n = 595 babies).

considering medium-risk babies. The author does not know of other methods for similar correction. Using uncorrected figures would have made non-significant the differences between tertiary services and the other maternity service level subgroups in relation to medium-risk babies as well as the difference between intermediate and secondary maternity services, regarding high-risk babies.

The analysis of utilization rates for different levels of care showed a different distribution of babies by sector level, according to risk level, by the different maternity services groups. Such a difference could have played a role in the observed differences in adequacy rates by care and risk levels among maternity services subgroups (secondary, intermediate and tertiary) (Tables 6 and 7).

Comparison of mortality rates for "treated" and "untreated" groups did not support the hypothesis that "untreated" groups have a higher mortality rate than the "treated" maternity service and sector subgroups. Some limitation in the analysis was posed by the occurrence of mortality rates related to small numbers and absent "untreated" groups. In fact, such small numbers of "untreated" subgroups, especially found in relation to secondary level procedures at tertiary maternity services, should be expected as an expression of the non-trial setting of the study, considering that the related process standards were nearly 100%. Considering that effective technologies have been selected and satisfactory process indicators were produced, three events could help to explain the situation whereby "treated" babies did not present better outcomes than "untreated" babies (or babies "treated" with lower level technologies): (a) selective allocation of babies generating "confounding by indication" in the analysis of the effect of a technology, within a similar care level and in the analysis of different levels of care; (b) unsatisfactory performance and low effectiveness; and (c) lack of precision in classifying risk level.

With respect to the hypothesis on the process—outcome relation that the relative risks of not receiving "treatment" when indicated were equal by sector and by maternity service level, results did not generally support that hypothesis.

The failure to confirm the above hypotheses—and more importantly, in the first hypothesis, the tendency for high and medium-risk babies to oppose it—should be related to the phenomenon of "confounding by indication" within the same risk level as a partial explanation,
unless the assumption of disastrous performance/effectiveness is considered for all study services.

In relation to the hypothesis according to which differences in mean adequacy rates were consistent with differences in general mortality rates between maternity services groups and care level sectors, a general trend to support it was observed. This tendency, albeit present, was neither universal nor statistically always significant as expected (Appendices 1 and 2), even considering $p < 0.10$. As satisfactory effectiveness was an assumption for the tendency to occur (Appendices 1 and 2), low and non-significant problems could be related to problems related to performance/low effectiveness.

Comparing now the different maternity care level subgroups, the advantage in terms of outcomes of tertiary and intermediate maternity services over secondary maternity services for high-risk babies (Table 3) could be related to tertiary and/or secondary level technologies. The significantly higher adequacy rate of tertiary care centres in comparison with that of secondary centres, for high-risk babies, in terms of secondary level technologies (Table 6), along with their inherent difference in relation to tertiary level technologies, supports the argument that either or both groups of technologies were responsible for tertiary care centres' advantage in terms of outcomes. These comparisons helped to expose another area of apparent contradiction: despite the fact that tertiary centres' overall outcomes were statistically more favourable than those obtained by intermediate centres only for medium-risk babies (Table 3), tertiary centres' overall adequacy rates were significantly higher than intermediate centres for all risk levels, for both secondary and tertiary technologies (Table 7). Considering the impact of tertiary technologies on the survival of high- and extreme-risk babies, these findings indicate low effectiveness for tertiary technologies available at tertiary centres. This hypothesis is reinforced by the observation that even compared to secondary care level services, tertiary centres did not differ significantly in outcomes for extreme-risk babies. On the other hand, the difference in outcomes (mortality rates) between intermediate care centres and secondary centres for high-risk babies ($p = 0.06$, Table 3) was not accompanied by a significant difference in adequacy rates for secondary level procedures, which could indicate higher effectiveness in performing secondary level technologies by intermediate centres compared to secondary centres (Tables 3 and 4, and 6). Another explanation for that finding—some degree of effectiveness for tertiary level procedures—is offset by the fact that differences in mortality rates between intermediate and secondary maternity services were less significant than differences between tertiary and secondary maternity services, considering the above indications for low effectiveness of tertiary technologies at tertiary maternity services.

Looking at sectors, the comparison of secondary sector mortality rates between both tertiary and intermediate maternity services subgroups, with the secondary maternity services subgroup, resulted in highly unfavourable rates for secondary services. These findings seem to indicate the influence of "confounding by indication", rather than a lack of effectiveness associated with secondary services, as mortality rates at tertiary and intermediate sectors were five times higher than at secondary sectors. The same phenomenon may have played a significant role in the difference between tertiary and intermediate maternity services in mortality rates for high-risk babies at the tertiary/intermediate sector, and for extreme-risk babies at the secondary sector, considering the percentage of utilization of the tertiary sectors by those services and the non-significant difference in mortality rates between those services for each risk level as a whole (taking both secondary and tertiary sectors together).

"Confounding by indication" probably occurred in the comparison of the effect of tertiary and intermediate maternity services, by sector, possibly, as a consequence of medium- and high-risk babies who presented more severe conditions (within a similar risk level) being selected to receive "treatment" (higher levels of care, that is, intermediate or tertiary sector), as opposed to babies with less severe conditions. The apparent selection of extreme-risk babies might have been more complex, considering that adequacy rates were very similar at the tertiary sector for high- and extreme-risk babies, while at the secondary sector extreme-risk babies received less care than high-risk babies for both tertiary and intermediate care centres. Possibly, babies with little chance of surviving at a certain level of care group failed to receive the same quality care as those with better prospects. These hypotheses do not dismiss the possibility of simple lack of care for some severe cases (including severe high-risk babies), as a complementary argument, taking into account, e.g. that the frequency and duration of the use of mechanical ventilation for babies with respiratory distress syndrome at the study services were greater than those reported in the literature [20].

Considering that relations schematically described in Appendix 2 have an assumption of "satisfactory performance", the above indications of low effectiveness related to care level procedures of certain maternity services groups make the inferences related to the occurrence of patient selection ("confounding by indication") within and among sectors less tenable. This conclusion exposes an important frailty of the study approach for validating outcomes through analysis of the process–outcome relation regarding care level sectors. The study approach was more useful for comparisons made at the level of maternity services, where "confounding by indication" does not operate, than at the level of sectors. The argument for the low effectiveness of secondary and tertiary technologies at the study services, raised by the present study, was confirmed by another approach of the process–outcome relation to validate outcomes, that included outcome standards, which could be developed for two of the procedures analysed by the present study.
REFERENCES


22. Min. Health Ontario/Advisory Centre on Reproductive Medical Care, A Regionalized System for Reproductive Medical Care in Ontario, manuscript, 1979.


APPENDIX 1

Elements for the analysis of the relation process-outcome. Considering the referent homogeneous in terms of severity for the same risk level (within the same care level sector or service as a whole)

<table>
<thead>
<tr>
<th>Referent (indications)</th>
<th>Outcomes</th>
<th>Relation process-outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T^*T$</td>
<td>$T^+T$</td>
</tr>
<tr>
<td>Rate of adequacy (%)</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Vary according to percentage of adequacy following efficacy standards for treated patients if performance of procedures is satisfactory.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increasing/differential adequacy does not change the relation between $RR(\frac{T}{T})S_1$ and $RR(\frac{T}{T})S_2$ if performance of procedures is satisfactory.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$*$T treated patients $T^*$ untreated patients.

$^\dagger$RR—relative risk of not receiving treatment when indicated. $S_1$ and $S_2$ represent services of the same care level.

APPENDIX 2

Elements for the analysis of the relation process-outcome. Considering the referent non-homogeneous in terms of severity for a same risk level and the occurrence of bias in the selection of patients for treatment (within the same care level sector or service as a whole)

<table>
<thead>
<tr>
<th>Referent (indications)</th>
<th>Outcomes</th>
<th>Relation process-outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T^*T$</td>
<td>$T^+T$</td>
</tr>
<tr>
<td>Rate of adequacy (%)</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Vary with performance of procedures and percentage of adequacy, not following efficacy parameters for treated patients, due to bias in the selection process. The rate of adequacy could influence the RR of not receiving treatment when indicated: if selection is biased towards the more severe patients, an increase in adequacy rates decreases mortality for both treated and untreated patients, whereas if selection is biased towards the less severe patients, an increase in adequacy rates increases the mortality rate for both treated and untreated patients. Regardless of the kind of bias, if the percentage of untreated patients decreases, total ($T^+T$) mortality rates decrease if performance is satisfactory.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$*$T treated patients $T^*$ untreated patients.

$^\dagger$RR—relative risk of not receiving treatment when indicated. $S_1$ and $S_2$ represent services of the same care level.