

The principal sources of drinking and household water for the population examined were the rivers on the banks of which most villages were situated. There is no doubt that, less frequently, water was also obtained from wells, creeks, swamps and lakes, but even where known and sampled, the proportionate contribution of these sources to the total intake could not be estimated. The analytical results based on 14 river water samples are set out in Table 6, page 285. For convenient interpretation of the data the villages were arranged according to their geographical location. Villages on the Wogupmeri and Karawari Rivers were listed in the order of increasing distance from the source of the river. (See map Figure 25, page 217 ).

Examination of Table 6 gave the impression of an appreciable decrease in the concentration of all three elements, downstream from the junction of the Wogupmeri and Wogupman Rivers. It seems most probable that this effect was caused by dilution of the Wogupmeri River by Wogupman water which, presumably, contained a lesser concentration of the elements analysed. (Unfortunately, no sample was available for analysis from the Wogupman River. See sub-section 3.3.2.5.). Naturally, no complete consistency can be expected when river water analyses are based on single samples and particularly not for rivers subject to frequent flooding. Therefore, one would not be surprised by somewhat low values at Sikaium which appear inconsistent with the trend claimed. However, this departure from the trend may be explained by dilution from a nearby creek which was sampled and found to contain 5 ppb fluoride, less than 1 ppb strontium and 0.15 ppm calcium, or by backflow from the Wogupman River if in flood at the time of sampling.

TABLE 6.

The fluoride, strontium and calcium  
content of river water at 14 villages.

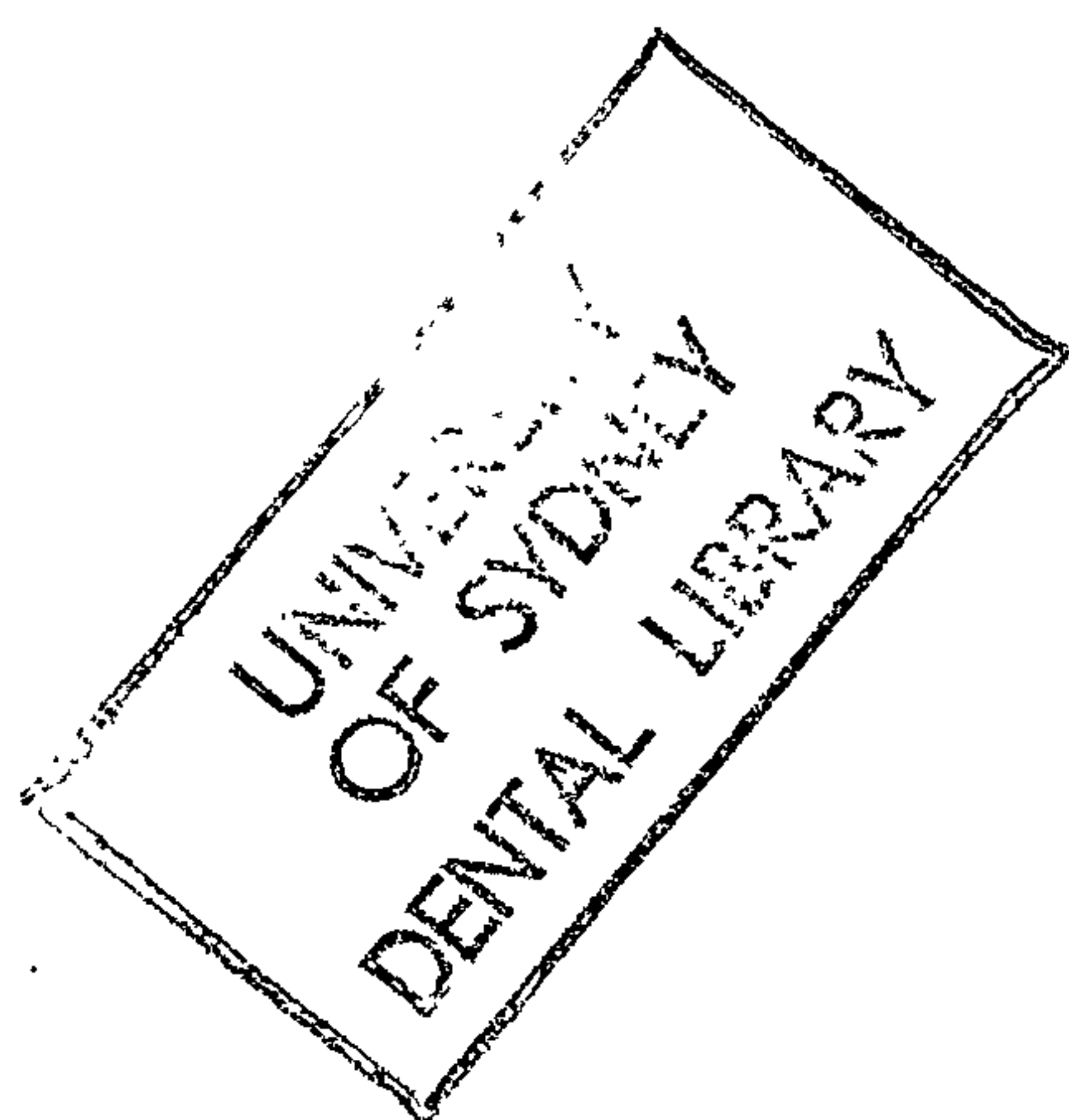
| Village |            | River      | Fluoride<br>(ppb) * | Strontium<br>(ppb) * | Calcium<br>(ppm) |
|---------|------------|------------|---------------------|----------------------|------------------|
| No.     | Name       |            |                     |                      |                  |
| 1       | Latoma     | Wogupmeri  | 24                  | 24                   | 10.0             |
| 2       | Danyig     | "          | 25                  | 18                   | 9.5              |
| 3       | Yenitabak  | "          | 23                  | 14                   | 7.6              |
| 5       | Barapidgin | "          | 27                  | 17                   | 6.9              |
| 4       | Sikaïum    | "          | 19                  | 8                    | 4.3              |
| 7       | Chimbut    | Karawari   | 17                  | 9                    | 5.2              |
| 8       | Maramba    | "          | 19                  | 11                   | 4.3              |
| 10      | Amongabi   | "          | 21                  | 9                    | 3.7              |
| 9       | Tangambut  | "          | 15                  | 10                   | 3.3              |
| 11      | Kundiman   | "          | 16                  | 6                    | 3.9              |
| 12      | Manjamai   | "          | 19                  | 14                   | 3.2              |
| 14      | Kabriman   | Blackwater | 11                  | 4                    | 0.9              |
| 15      | Mumeri     | Korosameri | 19                  | 19                   | 8.6              |
| 16      | Mindimbit  | Sepik      | 42                  | 33                   | 10.0             |

\* g/10<sup>9</sup> ml.

Further examination of Table 6 led the author to believe that there was a gradual decrease in the mineral content of the original Wogupmeri water (particularly noticeable in respect of calcium) with increasing distance from its source. The suspected gradient was by no means regular. However, considerable fluctuations would be expected because of changes in water levels and flow rates, dilutions or contributions made by tributary creeks, and perhaps other sources of variation. Thus, an apparent inconsistency, the relatively high fluoride and strontium levels at Manjamai, may reflect contributions from a creek which enters the Karawari River north of Kundiman Village and had 46 ppb fluoride and 34 ppb strontium in its water. The reason for the existence of a concentration gradient - if indeed it exists - is not clear to the author. However, it may be connected with the gradual settling out of clay particles onto which calcium and strontium are frequently adsorbed, or with complexing and settling out with organic matter. (436) It is not impossible that fluorides behave in a similar way. Fluoride is known to complex with aluminium, titanium, iron and silicon as well as with alum and other water coagulants. (437)

The mineral content of the Sepik River was considerably higher than than of the other rivers. Such differences could be expected because the Sepik River receives a very large number of tributaries over its 400 mile course, which arise in various geographical areas.

Comparison of the fluoride, strontium and calcium values obtained for a given water sample suggested that a strong correlation existed between the concentration of these elements. This relation-



ship was examined by regression analyses using the fourteen sets of values presented in Table 6. The product-moment correlation coefficients obtained for pairs of the elements are shown below.

|                    | n  | r    | p      |
|--------------------|----|------|--------|
| Strontium-Calcium  | 14 | 0.86 | <0.001 |
| Strontium-Fluoride | 14 | 0.88 | <0.001 |
| Calcium-Fluoride   | 14 | 0.74 | <0.001 |

The results indicated that, for all practical purposes, the concentration ratio of the three elements in water was a constant. As a matter of interest the fluoride : strontium ratio in surface enamel was examined, but no corresponding relationship could be demonstrated.

In absolute terms the strontium and calcium concentrations of water samples were rather low. This was apparent from Table 7, presented on page 288, where the results obtained in the present investigation were compared with results of studies carried out in various parts of the globe:

- 1) Barmes,<sup>(10)</sup> study of caries aetiology, wherein the mineral composition of water collected at 21 villages in the Sepik and Fly River areas was examined.<sup>(438)</sup>
- 2) A study of the public water supplies of the 100 largest cities of the United States,<sup>(439)</sup> which was in general agreement, as far as strontium and calcium concentration ranges were concerned, with the findings of Skougstad and Horr,<sup>(440)</sup> who analysed samples from 75 major rivers in the United States.

TABLE 7.

Comparison of the calcium, strontium and fluoride content of river water samples determined in the present investigation, with corresponding results obtained in three selected studies.

| Element           | Study                                     | Concentration |        |         |
|-------------------|-------------------------------------------|---------------|--------|---------|
|                   |                                           | Maximum       | Median | Minimum |
| Calcium<br>ppm    | Barmes, New Guinea <sup>(10)</sup> (438)  | 91            | 20     | 5       |
|                   | 100 U.S. cities. (439)                    | 145           | 26     | 0.0     |
|                   | Glass <u>et alii.</u> , Heliconia. (355)  | 200           | 6      | 2       |
|                   | Glass <u>et alii.</u> , Don Matias. (355) | 90            | 2      | 0       |
|                   | This investigation.                       | 10            | 4.7    | 0.9     |
| Strontium*<br>ppb | Barmes, New Guinea.                       | 850           | 150    | 50      |
|                   | 100 U.S. cities.                          | 1200          | 110    | 2.2     |
|                   | This investigation.                       | 33            | 13     | 4       |
| Fluoride*<br>ppb  | Barmes, New Guinea.                       | 160           | 50     | 20      |
|                   | 100 U.S. cities.                          | 7000          | 400    | 0.0     |
|                   | Glass <u>et alii.</u>                     | 300           | 100    | -       |
|                   | This investigation.                       | 42            | 19     | 11      |

\* Glass and associates<sup>(355)</sup> did not report detailed results for strontium; for fluoride the mean value was reported.

3) The study of Glass and associates, (355) who analysed the major and trace element content of 91 water samples collected from two Colombian villages, Heliconia and Don Matias, where a substantial difference was reported between the mean caries prevalence of the inhabitants of the two villages. (358)

The fluoride concentration of the river water samples ranged from 11 - 42 ppb. In dental literature, values of this order have been usually reported as "less than 0.1 ppm" (<100 ppb). Such statements carried the implication that variation below 100 ppb fluoride in the water had no practical significance, in so far as parameters associated with caries resistance or susceptibility were concerned. Previously observed similarities between the fluoride content of enamel and saliva in respect of certain villages, suggested to the author that this may not be true for the present sample. To examine this problem, the villages for which river water samples were available, were grouped according to the river on which they were situated. The mean fluoride content of the river water, surface enamel and saliva, as well as the mean caries experience, were calculated for these groups. The results are presented in Table 8, page 290. The results shown in the Table should be interpreted in full cognizance of the facts that river water analyses should be based on multiple samples which represent all seasonal variations and that this was not possible in the present investigation. Even to obtain a representative sample of a moving body of water at any one point in time is fraught with difficulties.

TABLE 8.

The mean concentrations of fluoride in river water samples collected at 14 villages, and the mean fluoride content of surface enamel, saliva and mean caries experience of subjects resident in those villages.

| Villages            | River            | Mean fluoride content |              |               | Caries Experience (DMFT) |
|---------------------|------------------|-----------------------|--------------|---------------|--------------------------|
|                     |                  | Water (ppb)           | Enamel (ppm) | Saliva (ppb)  |                          |
| 1, 2, 3, 4, 5       | Wogupmeri (N) *  | 23.6<br>(5)           | 625<br>(73)  | 27.6<br>(73)  | 5.2<br>(73)              |
| 7, 8, 9, 10, 11, 12 | Karawari (N) *   | 17.8<br>(6)           | 295<br>(131) | 14.6<br>(131) | 4.3<br>(131)             |
| 14                  | Blackwater (N) * | 11.0<br>(1)           | 211<br>(20)  | 15.3<br>(20)  | 0<br>(20)                |
| 15                  | Korosameri (N) * | 19.0<br>(1)           | 426<br>(13)  | 17.6<br>(13)  | 1.38<br>(13)             |
| 16                  | Sepik (N) *      | 42.0<br>(1)           | 555<br>(32)  | 21.6<br>(32)  | 1.09<br>(32)             |

\*(N) = Number of observations for each category.

However, the trends exhibited in Table 8 were too consistent to be disregarded, and were examined using regression and rank correlation analyses.

The correlation coefficients obtained for pairs of the variables included in Table 8 are shown in Table 9, page 292.

It was realized that the use of mean values reduces the variation in regression analysis, but this could not be avoided if the associations were to be examined on the basis of geographical entities (i.e. rivers), which seemed desirable. In a sense, averaging the water sample values tends to overcome deficiencies with regard to the representativeness of individual samples.

However, the same associations were also examined for the 14 villages for which river water samples were available. As expected, the correlation coefficients, presented in Table 10, page 293, were smaller, but in general agreement with the previous findings.

The results suggested the following conclusions:

- 1) In the present sample, variations in the mean fluoride level of river water were reflected consistently by the mean fluoride content of enamel. This fact could be established in spite of very low fluoride values in the water and without having taken into consideration other factors, such as surface wear, which were known to affect the fluoride content of enamel and were strongly operative in this population.

To the author's knowledge, there is no previous report in the literature of a corresponding association at similar fluoride levels.

TABLE 9.

Correlation coefficients between the mean fluoride content of water collected from five rivers, and the mean fluoride content of surface enamel and saliva, and mean caries experience of subjects resident in villages situated on those rivers.

| Pairs of variables          | N | Product-moment correlation |       | Rank correlation |      |
|-----------------------------|---|----------------------------|-------|------------------|------|
|                             |   | r                          | P     | r                | P    |
| F in water vs. F in enamel  | 5 | .72                        | <.05  | .9               | N/S* |
| F in water vs. F in saliva  | 5 | .51                        | N/S   | .8               | "    |
| F in water vs. DMFT         | 5 | -.002                      | N/S   | .3               | "    |
| F in enamel vs. DMFT        | 5 | .42                        | N/S   | .6               | "    |
| F in saliva vs. DMFT        | 5 | .47                        | N/S   | .3               | "    |
| F in enamel vs. F in saliva | 5 | .92                        | <.001 | .9               | "    |

\* When only five pairs of observations are available for rank correlation, no statistical significance can be demonstrated at the 5 per cent level if the r value is less than 1.0.

TABLE 10.

Correlation coefficients between the fluoride content of river water collected at 14 villages and the fluoride content of surface enamel and saliva, and caries experience of subjects resident in those villages.

| Pairs of variables          | N  | Product-moment correlation |       | Rank correlation |       |
|-----------------------------|----|----------------------------|-------|------------------|-------|
|                             |    | r                          | P     | r                | P     |
| F in water vs. F in enamel  | 14 | .41                        | <.10  | .66              | <.01  |
| F in water vs. F in saliva  | 14 | .51                        | <.05  | .80              | <.001 |
| F in water vs. DMFT         | 14 | -.08                       | N/S   | .20              | N/S   |
| F in enamel vs. DMFT        | 14 | .21                        | N/S   | .18              | N/S   |
| F in saliva vs. DMFT        | 14 | .01                        | N/S   | .13              | N/S   |
| F in enamel vs. F in saliva | 14 | .71                        | <.001 | .73              | <.001 |

2) A similar relationship appeared to exist in respect of the fluoride content of saliva. This observation is contrary to previously accepted findings<sup>(158)</sup> that only high (i.e. 3000 ppb = 3.0 ppm or more) water fluoride levels were reflected in the fluoride content of saliva.

3) It follows from the previous points that, on a group or village basis, there was a strong positive association between the fluoride content of enamel and the fluoride content of saliva. This relationship was confirmed on the basis of non-grouped individual results in the correlation matrix. (See Appendix, Part 2 a, b and c, pages 24 or 27, Variable 236 versus 243). The product-moment correlation coefficient,  $r$ , was 0.21,  $n = 298^*$ ,  $P < 0.001$ . However, because of the large number of pairs of observations involved, and the low proportion of the variation 'explained' at an  $r$  value of 0.21, little importance is attached to the significance level.

4) It was suspected and postulated on the basis of previously discussed results (sub-section 4.2.12.), that in the population examined, water-borne fluoride did not make a discernable contribution to freedom from caries. This contention was strongly supported by the results presented in Table 8, which showed that subjects in Village 13, all caries-free, had the lowest values for all forms of fluoride. Indeed, the relationship between caries experience and the fluoride content of water and

---

\* Through a computing error, the total number of observations appears as 301 on pages b 24 and 27. The actual number of observations was 298 as stated in the text.

saliva appeared to be indifferent, while a trend of positive association was noted between DMFT and the fluoride content of enamel. Examination of the correlation matrix showed that there was a weak positive association between the fluoride content of surface enamel (Variable 236) and all expressions of caries experience (Variables 10, 12, 14 and 229). Although the correlation coefficients were low for all combinations (0.11 - 0.13) the associations were significant at the 5 per cent level for 298 pairs of observations, despite the fact that only 1.2 - 1.7 per cent of variation in caries experience could be explained in terms of variation in the fluoride content of surface enamel. (See Appendix, Part 2 a, b and c, pages 13 and 27.)

Before interpreting this seemingly controversial finding three points need to be made:

a) The probability that the positive association demonstrated between the fluoride content of surface enamel and caries experience did not occur by chance was 20 : 1.

b) The existence of a correlation does not imply a cause and effect relationship between the associated variables. Even when a causal relationship exists, one variable is not necessarily the sole cause of the other.

c) Correlations can occur for various reasons, which include fortuity, interdependence, or the influence of a factor or factors on both variables to a similar extent in the same direction, or in opposite directions.<sup>(441)</sup> The actual reasons may or may not be apparent or explicable.

In the present case, a simple and logical, if not necessarily complete explanation can be offered.

It was shown that variations in the fluoride content of river waters, which are usually a good representation of environmental fluoride levels, were reflected in the fluoride content of surface enamel. However, with one exception, discussed below, the environmental fluoride levels appeared to have been too low throughout the area, to exert an appreciable effect on caries experience in competition with the effect of other determinants of caries susceptibility or resistance. In addition, the geographical distribution of the sample happened to be such, that large groups of caries-free subjects were exposed to low levels of environmental fluoride, while several of the villages where the mean caries prevalence was high, were situated in areas where the fluoride levels were somewhat higher.

The combination of these two circumstances is thought to have made a major contribution to the development of the positive association demonstrated between the fluoride content of enamel and caries experience. It is recognized, however, that other factors which were not apparent may have had additional influence.

An exception to the generally low environmental fluoride levels referred to in a preceding paragraph is thought to have occurred in the area of Villages 1, 2 and 3, where, in some subjects, signs of fluoride-rich enamel (sub-section 3.2.3.11.) and mild fluorosis were observed on clinical examination.

In previous surveys, (10) (13) (18) (19) (442) the author and others have consistently noted a pattern of gradual increase in

mean caries prevalence in villages, with increasing proximity to the source of rivers. This general pattern prevailed in the present survey with minor fluctuations, excepting the three villages nearest to the source of the Wogupmeri River, where caries prevalence was moderate compared with Villages 4 - 10, further downstream. This cannot be explained in terms of protection by water-borne fluoride alone, and it is thought that the comparatively low caries prevalence in these villages can be attributed to the availability of additional amounts of fluoride from other sources. This situation points to a need for extensive studies in an important but neglected field, the examination of total fluoride intake and utilization in man, and consideration of the effect on dental caries (and human metabolism in general) of fluoride obtained from sources other than water.

The mean strontium concentration of the water samples collected from each of the five rivers in the survey area were compared with the mean strontium content of enamel and mean caries experience of subjects resident in villages situated on those rivers. The results are presented in Table 11, page 298.

Examination of Table 11 suggested the presence of a strong trend of inverse association between the strontium content of river water and surface enamel. Associations between the variables shown in Table 11 were examined using regression and rank correlation analyses, based on the village group values and also according to individual villages for which river water and analyses were available. The correlation coefficients are shown in Tables 12 and 13, pages 299 and 300, respectively.

TABLE 11.

The mean concentration of strontium in water samples collected from five rivers, together with the mean strontium content of surface enamel, and mean caries experience of subjects resident in villages situated on those rivers.

| Villages       | River           | Mean strontium content |              | Caries Experience (DMFT) |
|----------------|-----------------|------------------------|--------------|--------------------------|
|                |                 | Water (ppb)            | Enamel (ppm) |                          |
| 1,2,3,4,5      | Wogupmeri (N)*  | 16.2<br>(5)            | 99<br>(73)   | 5.2<br>(73)              |
| 7,8,9,10,11,12 | Karawari (N)*   | 9.8<br>(6)             | 106<br>(131) | 4.3<br>(131)             |
| 14             | Blackwater (N)* | 4.0<br>(1)             | 113<br>(20)  | 0<br>(20)                |
| 15             | Korosameri (N)* | 19.0<br>(1)            | 78<br>(13)   | 1.38<br>(13)             |
| 16             | Sepik (N)*      | 33.0<br>(2)            | 74<br>(32)   | 1.09<br>(32)             |

\*(N) = Number of observations for each category.

TABLE 12.

Correlation coefficients between the mean strontium content of water collected from five rivers, the mean strontium content of surface enamel, and the mean caries experience of subjects resident in villages situated on those rivers.

| Pairs of variables           | N | Product-moment correlation |      | Rank correlation |      |
|------------------------------|---|----------------------------|------|------------------|------|
|                              |   | r                          | P    | r                | P    |
| Sr in water vs. Sr in enamel | 5 | -.91                       | <.01 | -1.00            | <.05 |
| Sr in water vs. DMFT         | 5 | -.079                      | N/S  | 0.10             | N/S  |
| Sr in enamel vs. DMFT        | 5 | .22                        | N/S  | -0.10            | N/S  |

TABLE 13.

Correlation coefficients between the strontium content of river water collected at 14 villages, the strontium content of surface enamel, and the caries experience of subjects resident in those villages.

| Pairs of variables           | N  | Product-moment correlation |      | Rank correlation |      |
|------------------------------|----|----------------------------|------|------------------|------|
|                              |    | r                          | P    | r                | P    |
| Sr in water vs. Sr in enamel | 14 | -.66                       | <.01 | -.59             | <.02 |
| Sr in water vs. DMFT         | 14 | -.26                       | N/S  | -.20             | N/S  |
| Sr in enamel vs. DMFT        | 14 | .54                        | <.02 | .56              | <.02 |

It seemed incredible that a strong and consistent inverse correlation should exist between the strontium content of water and enamel; nevertheless this was evident from both tables.

The following steps were taken in a search for an explanation of this unexpected result:

- 1) All analytical procedures were thoroughly re-traced. No errors were found outside the experimental errors reported, which were fully acceptable by usual analytical standards. This suggested that the results were valid and should not be rejected.

- 2) In view of the even distribution of strontium in enamel with depth,<sup>(149)</sup> and the recent finding by Tetteh,<sup>(443)</sup> that strontium taken up by surface enamel was readily released, it was considered unlikely that a significant amount of strontium would be taken up by the enamel permanently from the external environment after tooth eruption.

- 3) Consideration was given to the levels of strontium concentration found in water and in enamel. It was recalled that the strontium concentration in our water samples was near the lower limits of the very wide range found in various fresh-water samples (Table 7, page 288). In contrast, the mean strontium content of enamel was at about the middle of the 'normal' range. This suggested that sources other than drinking water may have made the most substantial contributions to the strontium content of enamel.

4) Underwood<sup>(444)</sup> stated that "....With most trace elements the food, and not the water or the atmosphere supplies an overwhelming proportion of total daily intakes by animals and man." Examination of the literature indicated that strontium in water made a relatively small contribution to total strontium intake. The amount of strontium ingested varies with so many factors (geological environment, age and level of food consumption, preference of foods, etc.) that very wide ranges of intake are possible. Tipton and associates<sup>(445)</sup> reported mean daily intakes of 1.37 and 1.24 mg for two adults. Bedford and associates<sup>(446)</sup> found the range of mean daily strontium intake of 0.67 to 3.57 mg for children of different ages. Murphy, Page and Watt<sup>(447)</sup> found that the amount of strontium ingested in a single school-lunch varied from .06 mg to 1.05 mg. Strontium supplied in the water was estimated as 17.8 per cent of the total intake by Hadjimarkos<sup>(448)</sup> based on the statistics of Schroeder.<sup>(449)</sup>

The relationship between the strontium supplied in the water and in foods for the population examined is directly pertinent. Unfortunately, results of food analyses will not be available for the present sample for some time. However, Barmes<sup>(10)</sup> reported that the strontium content of sago consumed by a comparable population in New Guinea ranged from 0.6 ppm to 3.3 ppm. The amount of sago consumed by an adult person per day was estimated as not less than 1000 g by Barmes.<sup>(10)</sup> Similar conclusions were reached following a survey of nutrition and diet in respect of the present sample.<sup>(450)</sup> This is equivalent to 0.6 to 3.3 mg of strontium ingested from sago per day. Taking the average strontium content of water as 15 ppb ( $\mu\text{g/l}$ ), and allowing for a daily con-

sumption of 3 litres in consideration of the hot climate, water would contribute 45  $\mu$ g to the daily intake.

Therefore, the strontium ingested in water would be no more than 13 per cent and could be as little as 1.4 per cent of the strontium ingested from a single food source, sago. Moreover, sago was the poorest source of strontium in the range of local food-stuffs analysed in Barmes's investigation,<sup>(10)</sup> taro and sweet potato contained about five and ten times as much respectively. Grasses grown on average type soils (with respect to strontium content) are known to contain up to 160 ppm strontium<sup>(451)</sup> and certain species of grasses and leaves form a substantial part of the daily diet of the population examined.<sup>(450)</sup> Accumulator plants may contain over 20,000 ppm strontium.<sup>(452)</sup>

Thus, it was concluded that, because an overwhelming proportion of the total strontium intake was most probably obtained from foods, there was no reason to expect that the strontium content of enamel would reflect the strontium content of water.

It would be more difficult to account for the inverse trend between the strontium content of water and enamel. The explanation, which as yet cannot be supported by direct evidence, may be connected with the strontium content of village soils. In a similar situation,<sup>(10)</sup> villages located on the Sepik River, and on the lower reaches of its tributaries where the river banks were low, had garden soils of high strontium content. The opposite was true for villages situated on the upper reaches of a tributary where the river banks were high. Barmes<sup>(10)</sup> attributed this to the

deposition of silt of high mineral content in the gardens of low-lying villages during regular yearly floods. Thus, inhabitants of villages on the lower reaches of the tributaries where the strontium content of river water is low, (unless the river is in flood or receives backflow from the Sepik), eat garden produce grown in soils of high strontium concentration.

For villages in swampy areas, where the strontium content of water appeared to be invariably the lowest, the situation is more complex. In many such villages visited by the author no soil suitable for planting existed. However, sago is always plentiful in swamps, and bartering of sago in return for garden produce, was common practice. On the basis of available information no approximation could be made of the relative contribution of foods obtained from various sources to the total strontium intake for these villages.

Superimposed on geological, topographical and broad nutritional variations, which are likely to affect people living in a given village in a similar way, there are many factors which may cause considerable individual variation in the strontium content of enamel. A recent review, concerned with the absorption of strontium,<sup>(453)</sup> indicated the complexities in that area and there are many others. It is known that the presence of calcium in the intestinal canal has a prominent inverse effect on the absorption of strontium<sup>(453)</sup> During betel-chewing massive amounts of lime (calcium oxide) are ingested, which, presumably contain significant amounts of strontium. (Some limestones contain as much as 4000 ppm strontium.)<sup>(454)</sup> Is the betel-chewing practice a source of extra

strontium, or does it result in blocking absorption of strontium ingested in food and water? Even if this were known, one would have to have information concerning the betel-chewing habit of the mother during pregnancy and lactation, details of possible betel feeding to the infant and of acquisition of the habit if this happens during early childhood, to make some approximation of the contribution made by this single variable to the amount of strontium incorporated in the dental tissues of the child. Thus, it is concluded that our present knowledge is insufficient to identify reasons for individual variations in the strontium content of enamel. Apart from some specific exceptions, this seems to be true for other tissues and for most trace elements.

#### 4.3. Correlation matrix.

Part 2a of the Appendix contains the entire correlation matrix, which shows the product-moment correlation coefficient relating each variable to every other variable in turn.

In examining the correlation coefficients one should keep in mind at least three factors:

- 1) The size of  $r$ , which may take any numerical value between 0 and 1, indicating no correlation or complete correlation respectively, between the two variables.
- 2) The sign of  $r$ , which indicates whether the correlation is a direct one or an inverse one.
- 3) The square of  $r$  gives the coefficient of determination, which is the proportion of the total variation of the dependent variable that can be 'explained' (in a statistical

sense) by variation in the values of the independent variable. This is true for two-variable and for multiple correlations.

Part 2b of the Appendix gives the actual number of observations that were available for each pair of variables. It will be seen that the correlation coefficients were not always based on the same number of observations. The reasons for this may have been that some subjects were not available or not suitable for certain tests, or that a technical breakdown occurred in the field, while on other occasions samples or results were rejected. As a rule, the loss of observations was small and the available observations were sufficient for significance testing in every case.

Part 2 c of the Appendix shows the pairs of variables for which statistically significant values of the correlation coefficient were obtained. The conventional system for indicating significance levels has been used, that is, the signs \*\*\*, \*\* and \* indicate  $P < 0.001$ ,  $P < 0.01$  and  $P < 0.05$  respectively. Where no indication was given, the result was not statistically significant.

Details concerning the location of specific components of the matrix are given in the preface to Part 2 of the Appendix.

It was pointed out in sub-section 3.4.3. that the primary purpose of this analysis was to indicate variables which were likely to contribute to the 'explanation' of variation in caries

experience. Some important limitations of this type of analytical system, which should inspire caution and restraint in its use and interpretation, were also discussed in the same sub-section.

Within its limitations, the correlation matrix proved to be a very useful analytical tool. Several references were made to the matrix in Section 4.2. dealing with the data summary. These demonstrated the manner in which the matrix was used for secondary purposes, for example, to obtain further information in respect of suspected trends found on examination of the data summary, and to compare the behaviour of grouped and individual data.

It would be superfluous to detail here in words what will be evident to the reader on examination of the actual figures given in the matrix. However, a few summary observations will be made.

As suspected, there was a high intercorrelation between DMFT, DMFS and the severity score of carious lesions. The high correlation between the first two parameters indicated a virtually linear increase in the number of carious surfaces with increasing number of carious teeth, although the actual differences in the mean scores were low (see sub-section 4.2.1.). Similarly, the severity score of carious lesions was a near-linear function of the number of carious teeth and surfaces.

The correlation between the former expressions of caries experience and the severity score/DMFS (Variable 229) was considerably weaker. There are several contributing factors which may explain this result. First, for caries-free subjects calculation of the severity score/DMFS value requires division by

zero. This automatically excludes consideration of caries-free subjects for this measure. Secondly, it is clearly possible to have a high correlation between DMFS and the severity score of carious lesions, and yet have an almost constant value for the severity score/DMFS rating. Consider the following artificial example:

| Subject | DMFS | Severity score<br>of carious lesions | Severity score/DMFS |
|---------|------|--------------------------------------|---------------------|
| 1       | 3    | 20                                   | 6.7                 |
| 2       | 5    | 30                                   | 6.0                 |
| 3       | 8    | 50                                   | 6.25                |

For data of this type it is clear that a high correlation could exist between DMFS and the severity score of carious lesions.

However, since the severity score/DMFS is virtually constant, the correlation between this measure and DMFS would be close to zero.

A more likely situation is one in which the severity score/DMFS value is subject to greater fluctuation than the severity score itself, for a given DMFS. An example of this is as follows:

| Subject | DMFS | Severity score<br>of carious lesions | Severity score/DMFS |
|---------|------|--------------------------------------|---------------------|
| 1       | 3    | 20                                   | 6.67                |
| 2       | 5    | 40                                   | 8.00                |
| 3       | 8    | 55                                   | 6.9                 |

In such a case, a reasonably high correlation coefficient between DMFT and the severity score of carious lesions would still exist, but that between the severity score/DMFS itself would be considerably smaller. In following analyses no further use was

made of severity score/DMFS. But, as a given severity score of, say, 40 may have arisen as a result of a different number of DMFS, the severity score/DMFS values would be different, and would represent the average intensity with which each attack has occurred. This measure merits closer examination and will be the subject of a separate investigation.

Examination of the correlation coefficients between any of the expressions of caries experience and the independent variables will show that the  $r$  values are generally low. This was expected in a caries aetiology study where the relevance of some of the independent variables is unknown, and even when relevant, each independent variable represents only one of a large number of active determinants of the disease.

Examination of the  $r$  values for all independent variables against DMFT as the dependent variable, showed that only seven of the former had  $r$  values of 0.20 or more. These variables and the respective  $r$  values are listed below:

| Variable No. | Designation (abbreviated)*                         | $r$  |
|--------------|----------------------------------------------------|------|
| 19           | Plaque quantity; UL2                               | 0.25 |
| 73           | Colony type 5 in UL2 plaque (%)                    | 0.27 |
| 97           | Colony type 5 in UL2 plaque (total no.x mg plaque) | 0.30 |
| 203          | Pooled plaque quantity (2)                         | 0.27 |
| 205          | M-D diameter of upper central incisor              | 0.20 |
| 208          | Number of Areca nuts used/day                      | 0.25 |
| 234          | Strontium ppm in enamel                            | 0.22 |

\* See Table 1 for full designation.

These correlations, and several others with marginally smaller  $r$  values, were significant at the  $P < 0.001$  level. Altogether, there were 30 independent variables for which statistical significance levels of  $P < 0.05$  or higher could be demonstrated. However, in terms of practical significance, it should be taken into consideration that an  $r$  value of 0.20 corresponds to only 4.0 per cent of the total variation that can be 'explained' in terms of variation in the value of the independent variable. Thus, in the practical sense, if it were acting on its own, none of these variables would be of material significance.

It was not intended, and it would not be possible, to draw far-reaching conclusions from the correlation matrix, particularly in terms of causal associations. However, the finding of many correlations, of a relatively low order, between the dependent and independent variables is in accordance with the assertion made in the introduction that in the absence of a dominating determinant the joint effect of a large number of variables needs to be examined to predict a significant amount of variation in caries experience.

The correlation matrix also contains, in a compact form, a vast amount of information concerning the interrelationships between the independent variables themselves. This alone is a fruitful source of ideas for the research worker, and may provide a considerable amount of information with regard to mechanisms of the caries process. Detailed examination of the interrelationships between the independent variables will be the subject of a

separate investigation.

#### 4.4. Stepwise multiple linear regression analyses.

The multiple linear regression analyses used in processing the results can be divided into two groups.

- a) Analyses carried out to aid with reaching decisions regarding the use of variables concerned with the streptococcal flora of plaque.
- b) Analyses carried out to test the hypothesis stated in Specific aim 2.

The results will be presented in corresponding sections.

##### a) Sorting analyses.

It has been previously stated (sub-section 3.4.3.) that a decision had to be made whether to use streptococcal counts and classifications based on pooled plaque, or corresponding data based on UL2 plaque in the analyses. Information was also desired on a further point: Would a single subset of the three expressions of colony types (per cent of all colony types, number per mg wet weight plaque, number x mg wet weight plaque) 'explain' as high a proportion of the variation in the dependent variables as the three expressions together?

Six multiple linear regression analyses (MLRA 1-6) were carried out to provide information on these points as dependent variables. All subjects for whom data were available were included in the analyses. The results were examined at the point

where the reduction in the error sum of squares was statistically significant at the  $P = 0.10$  level. The variables which had entered the analyses, the order in which they were selected by the analytical procedure, and the value of the multiple correlation coefficient,  $R$ , reached at that point are presented in Table 14, page 313.

Further details of the analyses, such as the stepwise increase in the values of the multiple correlation coefficient, changes in the error sum of squares and the  $F$  value at entering a variable are presented in Part 3 of the Appendix (MLRA 1-6).

For an alternative evaluation, the  $R$  values attained after 10 variables had been entered in the process were examined. The results are summarized below.

| Independent variable class                           | R values for dependent variables |                             |
|------------------------------------------------------|----------------------------------|-----------------------------|
|                                                      | DMFT                             | Severity of carious lesions |
| Streptococci in pooled plaque<br>(All 3 expressions) | 0.34                             | 0.32                        |
| Streptococci in UL2 plaque<br>(All 3 expressions)    | 0.45                             | 0.44                        |
| Streptococci in UL2 plaque<br>(No x mg used only)    | 0.39                             | 0.34                        |

Examination of the results showed that, the streptococcal counts and classifications based on the UL2 plaque explained a higher proportion of variation in caries experience than the same variables based on pooled plaque. The extent of the difference was considered sufficient to restrict further analyses to UL2 plaque data.

TABLE 14.

Correlation coefficients between two measures of caries experience and various expressions of pooled and UL2 plaque streptococci. (MLRA-s 1-6).

| MLRA  | Independent variable class.<br>(Number of subjects)                             | Dependent variable                |      |                             |      |
|-------|---------------------------------------------------------------------------------|-----------------------------------|------|-----------------------------|------|
|       |                                                                                 | DMFT                              |      | Severity of carious lesions |      |
|       |                                                                                 | Ind.variable entered              | R    | Ind.variable entered        | R    |
| 1 & 2 | Streptococci in pooled plaque<br>All 3 expressions used<br>(279)                | 31 Colony type 1 %*               |      | 35 Colony type 5 %          |      |
|       |                                                                                 | 35 " " 5 %                        |      | 56 " " 2 x                  |      |
|       |                                                                                 | 56 " " 2 x                        |      | 59 " " 5 x                  | 0.21 |
|       |                                                                                 | 59 " " 5 x                        |      |                             |      |
|       |                                                                                 | 62 " " 8 x                        |      |                             |      |
|       |                                                                                 | 29 Tot.No.Strep/mg                | 0.30 |                             |      |
| 3 & 4 | Streptococci in UL2 plaque<br>All 3 expressions used<br>(190)                   | 97 Colony type 5 x                |      | 97 Colony type 5 x          |      |
|       |                                                                                 | 85 " " 5 ÷                        |      | 70 " " 2 %                  |      |
|       |                                                                                 | 100 " " 8 x                       |      | 85 " " 5 ÷                  |      |
|       |                                                                                 | 73 " " 5 %                        | 0.39 | 94 " " 2 x                  |      |
|       |                                                                                 |                                   |      | 100 " " 8 x                 | 0.34 |
| 5 & 6 | Streptococci in UL2 plaque<br>Number of colonies x mg plaque used only<br>(190) | 97 Colony type 5 x<br>100 " " 8 x | 0.34 | 97 Colony type 5 x          | 0.26 |

\* % = as per cent (Variables 31 - 42 and 69 - 80, Table 1).

÷ = as number/mg plaque (Variables 43 - 54 and 81 - 92, Table 1).

x = as number x mg plaque (Variables 55 - 66 and 93 - 104, Table 1).

It was clear from the R values, that, a lesser proportion of variation in caries experience could be explained when only one expression of the colony types was included in the analyses, although the differences were not great. Therefore, the decision was made to allow all expressions of colony types to compete in subsequent analyses. It was noted that, of the three expressions of colony types, "number x mg plaque" featured more prominently than others.

Examination of variables entered in the analyses indicated that Colony type 5 had a dominant role in both pooled and UL2 plaque data. It is notable that when DMFT was the dependent variable, all three expressions of this colony type in UL2 plaque entered the analysis in competition for inclusion from all other colony types.

b) Testing the hypothesis that "sets of the intra-oral variables measured have a significant capacity to predict variation in individual caries experience".

The multiple linear regression analyses used to examine the hypothesis were carried out in several steps.

For the first step, the available independent variables were divided into three sets on the basis of their relevance to each other, and were submitted to the analytical process in separate sets against three expressions of caries experience, DMFT, DMFS and the severity score of carious lesions, as dependent variables.

It was pointed out earlier that one reason for close study of the correlation matrix was to limit the number of variables

which would be subjected to multiple regression analysis, by choosing those which appeared to be the most relevant to caries experience. However, the decision to use data associated with UL2 plaque and not pooled plaque, resulted in a considerable reduction in the number of variables and further reduction was unnecessary, because the computer programs available for multiple regression analyses were capable of handling the number of variables included in the sets. In addition, stepwise multiple regression analyses take account of the relative importance of the variables used. Therefore, no further limitations had to be imposed on the number of variables in constituting the sets.

Set 1: Streptococcal counts and classifications based on UL2 plaque, together with associated variables and variables of general relevance such as age (Variables 2, 16, 19, 22 - 26, 67 - 104, 201 and 203).

Set 2: Variables related to the chemical composition of enamel and saliva, together with associated variables and variables of general relevance (Variables 2, 7, 9, 16, 22, 201, 203, 205 - 209, 213 - 215, 217 - 220, 230 - 236, 239, 240 and 243).

Set 3: Variables related to the chemical composition of enamel, together with a reduced set of associated variables (Variables 2, 22, 205 - 209, 220, 230 - 236).

This approach was used for the following reasons:

- 1) To compare the capacity of the three sets of variables to 'explain' variation in caries experience.

- 2) To provide a basis for selection of variables from Sets

1 and 2, which contributed most to the explanation of variation in caries experience, for competition in the analyses performed in the second step.

3) To examine whether the results produced by one or more of the sets are capable of confirming or rejecting the hypothesis stated.

All subjects who had data in respect of all variables included in the sets, were considered in the analyses (MLRA 7 - 15). The analyses were allowed to progress, until the ten variables with the highest combined capacity to explain variation in caries experience had been entered in the process. Details of the analyses showing the ten variables in the order of selection, the stepwise increase in the values of the multiple correlation coefficient,  $R$ , the changes in the error sum of squares, as well as the  $F$  value at entering a variable are presented in Part 3 of the Appendix (MLRA 7 - 15).

The values obtained for the multiple correlation coefficient, the significance levels attained, and the number of subjects included in each analysis have been summarized in Table 15, page 317.

The significance levels shown in Table 15 were obtained using the formula given below and referring to tables of  $F$  distribution.

$$F_{p, n - p - 1} = \frac{R^2 (n - p - 1)}{(1 - R^2) p}$$

Where  $n$  = the number of subjects and  $p$  = the number of variables used to obtain the correlation coefficient,  $R$ .

TABLE 15.

Values of the multiple correlation coefficient and significance levels, for dependent variables representing caries experience and three sets of independent variables associated with properties of plaque (Set 1), enamel and saliva (Set 2) and a reduced set of variables (Set 3). (MLRA-s 7 - 15).

| MLRA     | Independent variables | Number of subjects | Dependent variables |       |       |       |            |       |
|----------|-----------------------|--------------------|---------------------|-------|-------|-------|------------|-------|
|          |                       |                    | DMFT                |       | DMFS  |       | Sev. Score |       |
|          |                       |                    | R                   | P<    | R     | P<    | R          | P<    |
| 7,8,9    | Set 1.                | 104                | 0.62*               | 0.001 | 0.62* | 0.001 | 0.59       | 0.001 |
| 10,11,12 | Set 2.                | 150                | 0.52                | 0.001 | 0.52  | 0.001 | 0.51       | 0.001 |
| 13,14,15 | Set 3.                | 178                | 0.42                | 0.001 | 0.40  | 0.001 | 0.40       | 0.001 |

\* The point where the reduction of the error sum of squares is not significant at the  $P < 0.10$  level had not been reached in these analyses.

Examination of Table 15 showed that Set 1, containing variables associated with the properties and streptococcal flora of plaque had the highest capacity to 'explain' variation for all three expressions of caries experience, although the corresponding capacity of Set 2 variables, associated with the composition and properties of enamel and saliva was far from negligible. Reference to MLRA outputs 7 - 9, Part 3, Appendix, will show that in Set 1 prominent roles were played by Colony types 5, 10 and 2.

Colony type 5 was the first, second and third independent variable to enter the analysis, when DMFT, DMFS and the severity score of carious lesions were the dependent variables respectively. Colony type 10 entered first and second versus DMFS and DMFT, respectively, while Colony type 2, in two forms, entered first and second when the severity score was the dependent variable. Pooled plaque quantity (2) also seemed important.

Analyses involving Set 2 were dominated by the strontium content of enamel and pooled plaque quantity (2), which were the first and second variables to enter the analyses versus all three dependent variables. (Pooled plaque quantity was included in both Sets 1 and 2, because of the possible influence of plaque on the composition of enamel.) For details see the Appendix, Part 3, MLRA outputs 10 - 12.

The reduction in the R values when the reduced Set 3 variables were used was, at this stage, attributed to the elimination of a

number of variables of importance, although the fact that this step was associated with an increase in the number of subjects included in the analysis was also noted. The variables excluded from Set 2 to form Set 3 included pooled plaque quantity (2), the viscosity, and the bicarbonate and calcium contents of saliva, all of which have played a material role in the analyses. The number of subjects included in the analyses increases with the reduction in the number of variables used, because the proportion of subjects who have no data for certain variables and are therefore excluded from the process, is also reduced. Thus, changes in the variables used also involve an involuntary 'selection' of subjects included in the analyses.

Examination of the R values and significance levels in Table 15, showed that all three sets of intra-oral variables measured had a significant capacity to explain variation in individual caries experience. Therefore, the hypothesis stated in Specific Aim 2 has been confirmed.

However, it was decided to proceed with the second step of the analyses, and examine the combined effect of the variables included in Sets 1 and 2. For this purpose the ten best independent variables were combined from each set, and analysed with DMFT as dependent variable (MLRA 16).

The analysis was allowed to proceed until the best 10 variables (of the 20 available) had been included in the process. The main features of the analysis are set out in Table 16, page 320, and the usual details are presented in the Appendix, Part 3, MLRA 16.

TABLE 16.

Values of the multiple correlation coefficient and significance levels, for caries experience (DMFT) as dependent variable and 10 selected independent variables representing properties of plaque, enamel and saliva. (MLRA 16).

| Indep. variable class                                                                               | Number of subjects | Dependent variable: DMFT (10) |       |        |
|-----------------------------------------------------------------------------------------------------|--------------------|-------------------------------|-------|--------|
|                                                                                                     |                    | Indep. variable entered       | R     | P      |
| Best 10 variables related to plaque<br><br>+<br><br>Best 10 variables related to enamel and saliva. | 157                | Strontium in enamel (234)     | 0.48* | <0.001 |
|                                                                                                     |                    | Calcium in saliva (240)       |       |        |
|                                                                                                     |                    | Pooled plaque quantity (203)  |       |        |
|                                                                                                     |                    | Bicarbonate in saliva (215)   |       |        |
|                                                                                                     |                    | Calcium in biopsy (220)       |       |        |
|                                                                                                     |                    | Colony type 5 (85)            | 0.52  | <0.001 |
|                                                                                                     |                    | M-D diam.central inc. (205)   |       |        |
|                                                                                                     |                    | Colony type 11 (79)           |       |        |
|                                                                                                     |                    | Colony type 3 (95)            |       |        |
|                                                                                                     |                    | Colony type 7 (75)            |       |        |

\* Past this point inclusion of further variables in the analysis did not produce a statistically significant reduction in the error sum of squares at the  $P < 0.10$  level.

The most important point to be noted on examination of Table 16 is that, the combination of the best ten variables selected from the two sets, produced an R value which was considerably smaller than the R value obtained by using Set 1 variables alone, and equal only to that attained when Set 2 variables were used in MLRA-s 7 - 15. As no error was found in the computing procedure, the result was attributed to the fact that the combination of the variables used resulted in the 'selection' of a sub-population which was different from those included in previous analyses. This in turn suggested the existence of some heterogeneity in the population which had not been recognized previously.

In considering this point further, it was assumed that in respect of susceptibility or resistance to caries, the major division was between subjects who had caries, regardless of the extent, and those who were caries-free. Such a dividing line cannot be absolute, because at least some of the subjects who have not yet succumbed to caries attack were likely to do so in the future. However, it seemed reasonable to argue that, restriction of the last analysis in this series to subjects who have experienced caries would improve the homogeneity of the sample, and a higher percentage of the variation in caries experience would be explained. If only carious subjects were to be included in the analysis, the severity of the disease becomes of increased importance. Therefore, it was decided to use the severity score of carious lesions as the dependent variable on the assumption that it may be a better measure of the true extent of the disease than the DMFT or DMFS scores.

In the selection of independent variables for this analysis it was kept in mind that when variables associated with plaque were used in previous analyses (Set 1 in MLRA 7 - 15), the point where the reduction in the error sum of squares was not significant at the  $P < 0.10$  level had not been reached at an R value of 0.62. This indicated further potential on the part of plaque-associated variables. The importance of variables related to the properties of saliva was shown by the considerable reduction in the R value when these variables were omitted from Set 2 to form Set 3 in MLRA-s 13 - 15, and again in MLRA 16. Of the trace elements in enamel only strontium featured in the analyses, but calculations showed no significant correlation between the strontium content of enamel and caries prevalence when only carious subjects were considered ( $r = 0.142$ ,  $n = 160$ , N/S). This finding suggested that, correlation between this variable and caries prevalence in the whole sample probably depended on associations between the complex distribution characteristics of strontium, previously discussed, and distribution characteristics of caries within the sample, rather than on actual relevance to the carious process.

The independent variables selected were: 2, 7, 9, 16, 19, 21 - 26, 67 - 104, 203, 205 - 209, 213 - 215 and 217 - 219. All subjects with caries who had data in respect of all these variables were included in the analysis. The principal features of the analysis are presented in Table 17, page 323. Further details can be seen in the Appendix, Part 3, MLRA 17.

Examination of Table 17 showed that an R value of 0.74 has been reached, and it is possible that the addition of more

TABLE 17.

Values of the multiple correlation coefficient and significance level attained in MLRA 17, for the severity score of carious lesions as dependent variable, and 10 selected independent variables, with caries-free subjects excluded from the analysis.

| Dependent variable: Severity score of carious lesions (14) |                                           |       |        |
|------------------------------------------------------------|-------------------------------------------|-------|--------|
| n                                                          | Independent variable entered              | R     | P      |
| 89                                                         | -Colony type 2 per cent of all types (70) |       |        |
|                                                            | Colony type 2 x mg plaque (94)            |       |        |
|                                                            | Colony type 5 x mg plaque (97)            |       |        |
|                                                            | Bicarbonate content of saliva (215)       |       |        |
|                                                            | Colony type 9 x mg plaque (101)           |       |        |
|                                                            | Plaque activity: resting pH (23)          |       |        |
|                                                            | Colony type 10 x mg plaque (102)          |       |        |
|                                                            | Viscosity of saliva (214)                 |       |        |
|                                                            | Colony type 11 per mg plaque (91)         |       |        |
|                                                            | Colony type 2 per mg plaque (82)          | 0.74* | <0.001 |

\* The point where the reduction of the error sum of squares is not significant at the  $P < 0.10$  level, had not been reached.

variables may have produced further significant improvement. The result is consistent with the contention that the previously found reduction in the R value (MLRA 7 compared with MLRA 16), reflected heterogeneity of the population. Moreover, the high R value attained in the last analysis, from which caries-free subjects were excluded, suggests that the division between complete freedom from caries and suffering from the disease to any extent is associated with a considerable reduction of the heterogeneity of the population, as far as the variables included in the analyses are concerned.

Several forms of heterogeneity may be recognized. The form indicated in the previous paragraph implies that the change of caries experience, which may take place for unit change in the independent variables (taken collectively) may not be the same over the whole range of caries experience. This implies a departure from linearity in the relationship between caries experience and the collective effect of the independent variables included in the analyses. The substantial increase in the R value, which followed the exclusion of the caries-free subjects from the analysis, may be an example of this situation; a major departure from linearity would occur if caries-free subjects were highly homogenous as a group, but would lie comparatively distant from the line of best fit for all subjects. The opposite of this proposition is also possible; caries-free subjects may be the most heterogenous group in respect of the variables examined and their removal from the analyses would therefore result in a material reduction in the error sum of squares, i.e. a material increase in the R value.

In contemplating a possible source of heterogeneity in this context, one should recognise that freedom from caries may be the consequence of high resistance to caries, or of environmental circumstances which do not predispose to the disease. If the variables measured reflect the oral environment, they could vary considerably between persons who are caries-free for one or the other reason. Data obtained in the present investigation can make a substantial contribution to the examination of this problem, which will form a separate study. This general subject will be discussed further in Section 4.5., dealing with discriminant function analyses.

Finally, heterogeneity of a completely different type occurs if some variables, which have not been measured, modify the effect of those included in the analyses, or influence caries independently, particularly if sections of the sample populations are affected to a different extent.

In examining the variables which contributed to the high R value attained in MLRA 17 one should recall that no matter how strong a correlation exists between variables, in biological systems cause-effect relationships can be proven only by reproduction of the results in an appropriate experimental model. However, there can be no objection to brief consideration of the possible relevance of these variables. In the light of present knowledge the relevance of streptococci in general cannot be doubted, although not to the exclusion of other organisms. In the present study, Colony types 2,5 and 10 appeared to be of importance in previous analyses, as well as in the present one.

This analysis has been dominated by Colony type 2, which appeared amongst the first ten variables selected in all three expressions. The fact that most colony types appeared as "number x mg plaque" seems to support the importance of plaque quantity, which featured prominently in other analyses. The bicarbonate content of saliva represents the foremost system contributing to its buffering capacity, while viscosity, a rather controversial property in this context, seems to be relevant in this population (see following paragraph). A low resting pH of plaque may indicate recent intake of fermentable carbohydrate, slow carbohydrate clearance, utilization of storage metabolites by plaque bacteria, or a combination of these factors, all of which may have aetiological significance.

It is of interest to note that, although subjects in the caries-free Village Group 4 had the highest mean viscosity of saliva, for subjects with caries, viscosity showed a direct association with caries experience. Thus, viscosity of saliva was one of the sources of heterogeneity between subjects who experienced caries attack and those who were caries-free. There was also a high degree of heterogeneity within caries free subjects in this respect, in that, at 67.7, the mean viscosity of saliva of 42 subjects resident in Villages 13 and 14, was significantly higher than that of 44 caries-free subjects resident in other villages, 50.9, ( $t = 3.03$ ,  $P < 0.01$ ).

In concluding this Section, aspects concerning the proportion of variation in caries experience 'explained' by the sets of independent variables used will be considered in the light of practicality.

It should be realized that independent variables may also have a capacity to 'predict' variation in the dependent variable. The predictive capacity of a set of independent variables is measured by the closeness with which the values of the independent variables substituted in the regression equation estimate the observed value of the dependent variable, for any given set of subjects. If this can be done with a good measure of precision, then it may be inferred that the independent variables, either directly, or indirectly, influence caries experience or are a consequence of it. This aspect was mentioned, because the correlation coefficient,  $R$ , also contains a conventional and comprehensive measure of the predictive capacity of independent variables. However, evaluation of the data in this context will be carried out in a separate investigation.

Disregarding the preparatory steps, the multiple correlation coefficients attained using different sets of independent variables, in combination with various measures of caries experience as dependent variables, ranged from 0.40 to 0.74. Consideration of these values, which correspond to the 'explanation' of 16.0 and 54.8 per cent of the variation in caries experience respectively, in terms of the independent variable, leads to the question: What is a reasonable or acceptable proportion of the variation 'explained'? As far as the author knows there is no study in existence which could be regarded as a precedent or used for comparison, but it is clear that both the population and the parameters examined have a bearing on the issue.

In most Western populations the substrate supply to the plaque flora usually dominates caries experience. Therefore, if

a numerical estimate of the challenge incurred through the diet can be made, it should be possible to show a strong association between that estimate and caries experience. The same is true of the protective effect of fluoride. A high correlation in such a case merely points to the presence of a variable, which, through its overwhelming effect, hides the true complexity of the operating mechanism. Conversely, if a dominant variable is present and is not taken into account, it is most unlikely that a good correlation could be established between another variable or variables measured, and caries experience.

In the present study, there was no evidence that known dominating variables would have contributed significantly to the existence of the major caries contrasts, or to the individual variation of caries experience in the villages where high caries prevalence was found; thus the true complexity of the situation emerged.

In the purely statistical sense, a proportion of the variation 'explained', which satisfies the appropriate tests of significance is clearly acceptable. In the present study all of the main analyses complied with this requirement, yet two of them 'explained' only 16 per cent of the total variation in caries experience.

From the practical point of view, the following factors should be considered, all of which tend to reduce the proportion of the total variation that can be accounted for:

- 1) Not all variables that have an influence on caries

susceptibility or resistance have been measured. Even if one assumed that some of the variables for which data are available reflect others which were not measured, the coverage of causally-related variables could not have been complete.

2) Sampling or measurements taken at any one point in time, are not necessarily representative of the total past experience of the individual in respect of the variables concerned; these data are used to 'explain' caries scores which represent total past caries experience.

3) The extent to which certain tests or measurements reflect biological reality is questionable, and may vary from person to person.

4) All measurements and tests are associated with experimental error.

5) Undetected or undetectable heterogeneity may exist within the sample.

6) In multiple linear regression analysis the analytical procedure assumes a linear relationship between the dependent and independent variables. In fact the relationship could be of any form.

All the factors listed above combine to reduce the proportion of total variation explained by the independent variables to below what is theoretically possible. Although longitudinal observations could produce substantial improvements, because of the inherent deficiencies of the sampling, technical and statis-

tical procedures, one would never hope to 'explain' 100 per cent of the variation. It is conjectural how far short of the 100 per cent one would fall, but in this context, up to 55 per cent is considered a very substantial proportion of the variation that can be 'explained' in practice.

#### 4.5. Discriminant function analyses.

It will be recalled that, the purpose of conducting discriminant function analyses was to examine the extent to which individuals with materially different caries experience can be classified as such, on the basis of the information contained in the independent variables alone.

An important principle involved in the design of this type of analysis is that the subjects included in the procedure, should represent clearly different levels of the characteristic to be separated, in this case caries experience. The definition of the term 'clearly different levels' is a matter of interpretation, at the discretion of the investigator. There was no published information available concerning the likely discriminatory capacity of the analytical system when applied to a similar problem. It was reasoned that if separation of subjects with differing caries experience were at all possible, this would be most likely to succeed if subjects included in the analyses exhibited the highest possible contrast in respect of caries experience. At the same time, it was recognised that this decision will have consequences as far as the number of available observations were concerned.

In the final outcome, it was decided to attempt to separate subjects with six and more DMFT from those who were caries-free, because this division provided the maximum contrast consistent with a substantial number of subjects having been included in the analyses. It should be noted that it was not necessary to use more than one measure of caries experience, because the same subjects would have been included in the analyses and the results would have been the same if DMFS or the severity score of carious lesions had been used.

With respect to the selection of the independent variables on which the classification of subjects would depend, it was decided to employ Sets 1 and 2 used in the multiple linear regression analyses, for reasons previously explained (MLRA 7 - 15, see pages 315 and 316).

All subjects with six or more DMFT and all caries-free subjects who had data for all variables included in the respective sets, were included in the analyses.

Thus, two analyses were carried out, using Set 1 and Set 2 variables respectively, to classify subjects who had either six or more, or no carious teeth, on the sole basis of the information contained in the variables included in those sets. These analyses will be referred to as DFA 1 and DFA 2 respectively.

For each of the analyses two figures are presented (Appendix, Part 4) which illustrate the rate at which individuals were misclassified on the basis of the value of their discriminant function.

Figure 1 shows the two types of misclassification which are labelled False Positive % and False Negative %. These errors may occur:

- a) because of the value of the discriminant function for a particular individual, he is classified as belonging to the  $DMFT \geq 6$  group, when it is known from the dental examination that he has 0 DMFT;
- b) conversely, because of his discriminant function value, a subject is classified as belonging to the 0 DMFT group, when in fact the dental examination showed that he has six or more DMFT.

The symmetry of the curves in relation to the axes (indicating a similar distribution of the discriminant function values for the two groups) makes it irrelevant to attempt to identify an axis as corresponding to a particular kind of error.

However, the rate of misclassification also depends upon the critical value of the discriminant function chosen for the purpose of allocating subjects to one group or another.

There is an argument which suggests that accepting equality for the two kinds of error is not necessarily the best decision in all cases. It is clear that a subject in the group for which DMFT is six or more, whose discriminant function value indicates that he should belong to the 0 DMFT group, has been misclassified by this process. This is, in fact, a true error.

With respect to the possibility of a "false error" one should consider the case of two caries-free young adults in the

present population.

One of these subjects, who lives in a village where a high level of caries occurs, is in a sense a "true" caries-free person. The reason for this is that in a "high caries" village the challenge in respect of caries has been obviously presented by the environment, but our subject has resisted it. Thus, unless he were very different from his peers in some way, and as a consequence he were exempted from the environmental challenge, one should accept that his freedom from caries is vested in his resistance.

In respect of the other subject, who lives in a village where caries does not occur in any individual, it is possible that no challenge has been presented and, given that it were, he would succumb. Thus, a caries-free person from a caries-free village may be one whose resistance or susceptibility to caries (as reflected by the variables measured) may not differ from a person with caries experience. Classification of such a person as one who belongs to the high caries group may not be a true error.

Therefore, it could be argued that the value of the discriminant function at which one decides to which group the subjects should belong, should be such that a large error rate is tolerated for the true kind of error and a small error rate for the other kind.

Alternatively, if one is willing to accept the proposition that errors of both kinds have been made, and that discrimination is incomplete because some variables of significance have not been measured, a dividing line which is such that the errors of mis-

classification are equal, would seem to be the most appropriate. In this work equal errors of misclassification will be used.

In the diagram shown in Figure 1, Part 4, Appendix, the further the curve is removed from the hypotenuse of the triangle, the better is the discrimination. The magnitude of the error of misclassification, when these errors are assumed to be equal, can be estimated from the diagram. As an approximation, the percentage misclassification corresponds to the points on each axis, which are perpendicular to the point on the curve which is furthest from the hypotenuse.

Figure 2 shows the distribution of subjects for each discriminant function analysis. The interpretation of these figures will be explained with reference to Figure 2, DFA 1. Caries-free subjects (designated Group 1 in this computer output), are represented by the sign  $\square$  while the sign  $\times$  represents subjects with six or more carious teeth (Group 2). An individual is not necessarily represented by the same number of signs in the two groups; nor is a subject represented by an integral number of signs unless the total number of subjects in a group is divisible by a convenient number to give an integer. In Figure 2, DFA 1, 4.87  $\square$  signs represent one caries free individual, while 9.77  $\times$  signs are equivalent to a person with six or more carious teeth. However, 5  $\square$  signs and 10  $\times$  signs respectively, are sufficiently close approximations.

The discriminant function value, which corresponds most closely with the mean discriminant function of the subject or

subjects represented by a column of signs, is set out next to each column. Overprints of signs indicate instances when subjects who, on the basis of their actual caries scores, belong to different caries prevalence levels, have identical or closely comparable discriminant function values. Thus, overprints invariably represent cases of misclassification.

Table 18, page 336, shows in a summary form the result of the analyses, based on two sets of independent variables. For each analysis, Table 18 shows the number of subjects included, the mean discriminant function value for each caries prevalence group, the critical value of discriminant function on which the classification of individuals depends (based on equal errors of misclassification), and the number and percentage of individuals misclassified.

It is pointed out that the magnitude or sign of the mean discriminant function for the groups is of no importance in the context discussed. However, the difference between the mean values for Groups 1 and 2 represents the standardised distance between the groups. (See Section 3.4.5.)

Examination of Figures 1 and 2, DFA 1, Part 4, Appendix, will show that the separation of caries-free subjects from those with six or more carious teeth was remarkably good on the basis of the information contained in the variables included in Set 1. It is pointed out here that the graphical representation of the misclassification rate shows it to be marginally higher than it was in actual fact. The reason for this is not clear. The

TABLE 18.

Summary of results of discriminant function analyses. Classification of subjects with known caries experience (0 and 6 or more DMFT in Groups 1 and 2 respectively), using information contained in two sets of independent variables, representing properties of plaque (Set 1), and of enamel and saliva (Set 2).

| Variable<br>class | No. of subj. included<br>in the analyses |         |       | Discriminant function |         |                   | Misclassification |               |
|-------------------|------------------------------------------|---------|-------|-----------------------|---------|-------------------|-------------------|---------------|
|                   |                                          |         |       | Mean                  |         | Critical<br>value | No. of<br>subj.   | % of<br>subj. |
|                   | Group 1                                  | Group 2 | Total | Group 1               | Group 2 |                   |                   |               |
| Set 1.            | 62                                       | 31      | 93    | 90.28                 | 77.53   | 83.9              | 4                 | 4.3           |
| Set 2.            | 77                                       | 51      | 128   | -3.49                 | -6.71   | -5.15             | 23                | 18.0          |

overprints appearing in Figure 2 indicate that two caries-free subjects plotted at values 78.92 and 80.75 were classified as if they had six or more carious teeth. In view of the argument presented earlier in this sub-section, this may have been a 'false error'. The critical discriminant function value for this analysis was 83.90 on the basis of equal errors of misclassification. Thus, a further two subjects, plotted at values 84.41 and 84.87 should be regarded as having been misclassified, although they do not overlap the main body of caries-free subjects. As the latter two subjects had six or more carious teeth, there is no doubt that if the judgement of using equal error rates was correct, the misclassification of these two subjects represents a 'true error'. Therefore, Table 18 shows that four subjects were misclassified. However, it would be unduly conservative not to point out the possibility that, if the first two misclassifications were indeed 'false errors', and the decision to use equal errors of misclassification proved to be wrong on further use of these analyses, one could claim that the discrimination between the two groups was absolute.

The analysis using Set 2 variables resulted in a higher rate of misclassification at 18.0 per cent. In the context of the complexity of caries aetiology, even this result is considered exceedingly good. Once again, a minor departure from the self-imposed rule of equal error rate, that is, taking 5.35 as the critical discriminant function value instead of 5.15 would have resulted in a reduction of the misclassification to 14.8 per cent.

Thus, it is concluded in respect of the third specific aim of this study that, individuals with materially different caries experience could be classified as such, on the basis of the information contained in two specific sets of independent variables, with error rates of 4.3 or 18.0 per cent, or less, respectively.

The outcome of these exploratory analyses was most promising. In future work it is planned to extend the use of discriminant function analysis along the lines employed in the present study, and to explore possible further applications using the data collected in this investigation as material to establish a model which may have universal application. A unique feature of discriminant function analysis is, that it takes into account contrasts between variables. Examination of the discriminant function for groups of subjects with contrasting caries prevalence could indicate the relative importance of variables, groups of variables or contrasts between variables which make material contribution to the separation of the groups. It would be of considerable interest to examine the power of the method to separate groups of subjects with progressively lesser contrasts in caries experience, and to determine whether the same variables are implicated at all levels. On an individual basis, examination of the data for subjects who were misclassified may disclose the features in which they differed most from other subjects at a similar caries prevalence level.

Like other statistical methods, discriminant function

analysis suffers from the disability that it can never provide proof of causal relationships. However, the method seems to hold considerable promise that the information it may yield regarding relationships between determinants of caries and the incidence or prevalence of the disease, may be more fundamental than that provided by commonly used statistical procedures.

CHAPTER 5.CONCLUSIONS AND SUMMARYUNIVERSITY  
OF SYDNEY  
DENTAL LIBRARY5.1. Precis of study.

In the broadest sense, this study was designed to examine the feasibility and value of the multi-factorial approach to the investigation of the aetiology of dental caries in man.

The specific aims of the study, detailed in sub-section 1.2.2., required:

- 1) Collection of quantitative data for selected intra-oral variables from individuals in a primitive population.
- 2) Examination of the capacity of sets of the variables to 'explain' variation in caries experience.
- 3) Investigation of the extent to which subjects with materially different levels of caries experience can be classified as such, on the sole basis of the information contained in the independent variables.

An extensive and informative body of data has been collected from a human sample of 301 subjects, aged 12 - 24 years. The sample was fully representative of the age stratum of the primitive population in New Guinea from which it was selected.

The data collected from each individual represented 114 intra-oral variables, and included four expressions of caries experience, selected characteristics of plaque, enamel and saliva

and aspects of dental morphology. Analyses of river water samples were carried out for fluoride, strontium and calcium content. In addition, the social, economic and environmental background of the population was investigated, with particular emphasis on diet and eating habits.

The observations were tabulated, and associations between caries experience and the independent variables were examined, using a data processing system developed for this purpose.

## 5.2. Conclusions.

To benefit from encompassing a broad spectrum of known or suspected determinants of caries, a large number and variety of intra-oral variables were included in the investigation. Several of the variables were examined for the first time, and many involved analytical work using micro-samples. For these reasons, demands for method development and application were unusually high.

Detailed descriptions given in the text, of over twenty methods and procedures developed or modified specifically for use in this investigation, involving the disciplines of epidemiology, analytical chemistry, biochemistry, physics, microbiology and data processing, show that demands in the area of methodology have been met.

It has been demonstrated for the first time, that flameless atomic absorption spectroscopy is a suitable method for the determination of trace metal constituents of surface enamel. Determinations of potassium, strontium and zinc were made at

high precision levels, using minute samples of surface enamel obtained by individual biopsies in-vivo. Semi-quantitative determination of barium was also possible.

Extension of the application of flameless atomic absorption spectroscopy to the determination of the trace metal composition of other oral biological materials and particularly to that of dental plaque, for which no such information is available at the epidemiological level, should be possible and would be desirable.

The difficulties concerning design, methodology and logistics usually encountered in this type of research were considerably increased by the setting of the study in a primitive and remote environment. However, it has been demonstrated that it is feasible to conduct investigations involving delicate techniques, even if the field work has to be accomplished under the most unfavourable conditions.

In this type of study, the caries pattern exhibited by the sample is of critical importance, because a sufficient range and contrast in caries is necessary for establishing associations between the parameters studied and caries experience.

In the population investigated, the caries experience of individuals ranged from 0 - 22 DMFT. The range of mean DMFT per person per village was 0.0 - 8.7. Corresponding contrasts were observed in the severity of carious lesions. Thus, the sample selected did provide the range and contrast in caries experience required for the investigation.

Collectively, the distribution pattern, wide range and high contrast found in caries prevalence is believed to be exceedingly rare, if not unique, for a racially homogenous primitive population, resident in villages situated in close proximity to each other, on a single river system.

The sharp contrast in caries prevalence could not be attributed either to differences in dietary patterns, or to variations in the level of environmental fluoride, which are the only two factors presently known to exert a decisive influence over the prevalence of caries at the population level. The main reasons for this conclusion can be summarised as follows:

The staple food and principal source of carbohydrate was sago for the whole population. No differences were detected in the diet, food preparation or frequency of meals that could account for the high contrasts in caries prevalence, either by a nutrition - diet survey or by the author and associates during numerous field visits.

It has been ascertained that the consumption of refined Western-type foods was either negligible or non-existent in this population. Moreover, there was an inverse trend between access to processed foods and caries experience.

The fluoride levels in river water were very low throughout the survey area (0.01 - 0.04 ppm). This was reflected in the low fluoride contents of surface enamel and saliva of subjects in most villages. (Sample means were 415 ppm and 19 ppb, respectively.)

Overall, the fluoride content of household water, surface enamel and saliva were either comparable for caries-free and high caries groups, or were slightly higher for the latter.

It was also demonstrated that the mean fluoride contents of both enamel and saliva were significantly different in the two villages where no caries was detected.

These observations suggest that for the sample as a whole, 1) freedom from caries, 2) very low prevalence and severity of carious lesions (1.1 - 2.2 mean DMFT per person per village) and 3) major contrasts in caries prevalence (0.0 - 2.2 mean DMFT per person per village, versus 5.2 - 8.7 mean DMFT per person per village), were essentially independent of environmental fluoride.

However, the moderate caries prevalence in the three villages situated nearest to the source of the Wogupmeri River (3.5 mean DMFT per person in contrast to 6.5 - 8.7 mean DMFT in neighbouring villages), was associated with generally higher fluoride content of surface enamel and saliva and in some subjects with clinically detectable signs of mild fluorosis. The moderate caries prevalence in these villages is attributed to the protective effect of fluoride, probably operating through a vehicle other than water, being superimposed on a combination of environmental influences which otherwise predispose to high caries experience.

The latter observations suggest that:

- 1) If the main caries contrast (i.e. between the high

caries group and the low and zero caries groups), had been due to the effect of fluoride, it would have been detected.

2) Fluoride would confer significant protection against caries in this population, as elsewhere, if it were available at adequate levels.

3) Where the fluoride content of water is considerably below optimal levels, the intake of fluoride from sources other than water assumes fundamental importance. Therefore, total fluoride intake warrants extensive investigation.

It would have been pointless to collect a large body of information without establishing a system of analysis which could be used for orderly statistical interpretation of the data. The system developed for this purpose in the present study has been shown to have that capability.

It is appropriate to make further conclusions with reference to the four major components of the system used to evaluate the information: the data summary, the correlation matrix, the stepwise multiple linear regression analyses and the discriminant function analyses.

The data summary contains condensed information regarding the 114 intra-oral variables examined, and, as far as the author knows, represents the most extensive published collection of parameters concerned with the oral environment of a defined human population.

Conclusions reached in respect of each variable or groups

of variables have been presented in the text and will not be detailed here. However, some general aspects need to be mentioned.

For several of the variables tabulated in the data summary, no comparable information exists in the literature. Where comparisons were possible, the mean values and ranges of the variables measured were in general agreement with other findings.

An exception is the calcium content of saliva, which was higher than that usually observed in Western populations. It ~~was shown that the elevated values occurred as a consequence of~~ betel-chewing.

The range of calcium : phosphorus ratios in surface enamel was remarkably narrow (mean 2.14, S.D. =  $\pm 0.07$ ). This indicated that, in the 299 subjects examined, very little true individual variation existed and suggested that in sound enamel, sampled to a uniform depth, variation in the Ca : P ratio may be much less than was previously assumed.

In practice, the summary tables have proved to be an invaluable source of reference. Apart from the primary purposes of showing mean values and variations, serving as a basis for comparison of values within and outside the sample, and indicating trends of associations, scrutiny of the data summary led to the detection of unusual, unexpected or aberrant values in several instances. This in turn re-directed attention to individuals or to details and sections of the data which required close personal examination or special treatment. With the presentation of the data summary, the first specific aim of the investigation has been accomplished.

The correlation matrix was used as a "pathfinder" much more extensively than is apparent from the text but by no means to its limit. The relationships between dependent and independent variables received close attention for each pair. Although examination of interrelationships between the independent variables was outside the scope of this study, it would have been difficult to resist following at least some ramifications of the work in this direction. Brief examination of selected aspects showed that interrelationships between independent variables were capable of logical interpretation and merit further study in detail.

The correlation matrix presented is probably the most comprehensive one in existence concerned with intra-oral variables relevant to the aetiology of dental caries.

Stepwise multiple linear regression analyses were carried out as sorting procedures and to examine whether it was possible to 'explain' a significant proportion of variation in caries experience by variation in the values of the independent variables.

When sets of variables, formed according to their relevance to broad aetiological areas, were submitted to the analyses, those associated with plaque and its microbial flora invariably had a better capacity to 'explain' variation in caries experience than variables representing the composition and properties of enamel and saliva. The respective multiple correlation coefficients, using sets of ten variables, were 0.62 and 0.52, each being significant at the  $P < 0.001$  level.

Amongst variables associated with plaque, the quantity of plaque and three colony types of streptococci (Colony types 2, 5 and 10) emerged as the most consistently important parameters.

Limited tests carried out using representative isolates led to the presumptive identification of about one half of the isolates in Colony category 2 as S. mutans, although on primary isolation the colony morphology was not at all typical. The identity of the remaining half of the isolates is uncertain, but they resemble S. mutans more closely than any other species. Colony type 10 had the recognised characteristics of S. mitis. Colony type 5 could not be classed into any of the established taxonomical categories, but it resembled S. sanguis to a limited extent.

Owing to their importance in the analyses, comprehensive identification of all colony types isolated shall be included amongst numerous follow-up studies already suggested in this text.

It is anticipated that some colony types represent more than one strain and possibly more than one Streptococcus species. Therefore, it is likely that the correlations demonstrated between various expressions of colony types and caries experience reflect the combined effect of a number of strains classified within that colony category. For any specific strain, the association may be stronger or weaker than it appears from the results.

The strong associations found in these analyses suggest that careful classification on the basis of colony morphology is

an effective means for screening viable streptococci present in plaque, but morphology should not be accepted as the sole criterion for establishing their identity.

In view of the results of presumptive identification and in cognizance with the uncertainties surrounding the taxonomy of S. mutans, the author believes that, a material proportion of isolates classified as Colony type 2 was, what is conventionally regarded as S. mutans.

The presence of S. mutans in the plaque samples was also evident from the findings of Bratthall, who examined aliquots of the same specimens using the fluorescent antibody method and detected the organism in 50 per cent of the samples (sub-section 4.2.6.).

According to the former criteria, S. mutans has been isolated in all villages and has been prevalent at all levels of caries experience, although it rarely amounted to more than 5 per cent of all streptococci isolated.

The fact that S. mutans has been isolated from all villages, suggests that the high intake of refined sucrose usually associated with Western-type diets is not a necessary requirement for the propagation and establishment of this organism in human dental plaque.

The strong associations demonstrated between caries experience and the proportions and absolute numbers of Streptococcus species in plaque, represented by Colony types 2, 5 and 10, probably in this order of (statistical) importance, suggest

further conclusions:

The prominence of Colony type 2 in the analyses is in accordance with claims made in respect of the aetiological importance of S. mutans. However, the demonstration of significant associations between caries experience and other species of streptococci, highlights the oversights that can occur if attention is focussed on a single species to the exclusion of other members of the plaque flora.

It has been shown that the acid producing capacity of plaque in the primitive population examined was at least equal to and possibly greater than that in Western societies. Conversely, the proportion of S. mutans to other streptococci in plaque was lower than the proportion usually found in Western populations. This suggests that S. mutans is not an indispensable contributor to the acid producing capacity of plaque.

Of the variables associated with the composition of surface enamel, barium, fluoride, magnesium, potassium and zinc played no material role in the analyses. However, a significant positive association was found between caries experience and the strontium content of surface enamel when the whole sample was considered. It has been contended by others that strontium is a 'beneficial' element, mainly because of inverse associations found between the strontium content of water and caries prevalence in epidemiological studies. The results of this investigation did not support this contention.

Moreover, evidence has been presented that in the population

examined, strontium ingested in the water makes a minor or negligible contribution to the total strontium intake.

It is not suggested, however, that the direct association demonstrated between the strontium concentration of enamel and caries experience has an aetiological significance. The strontium content of enamel may be merely a reflection of the availability of this element at the site during enamel formation. This in turn depends partly on the presence of strontium in the environment, and partly on factors which influence its absorption and utilization. Of the latter, competition with calcium at the intestinal absorption level seems to be particularly relevant in the present population. Until more information is forthcoming concerning these factors it is unlikely that the reasons for variations in the strontium content of enamel can be further elucidated.

Amongst the properties of saliva examined, the calcium content appeared to be of consistent importance in the analyses. The inverse association between this variable and caries experience is thought to have occurred as a consequence of betel-chewing. In the present investigation the examination of the relationship between betel-chewing and caries experience has been of an exploratory nature. However, available information suggests that betel-chewing may confer a degree of protection against caries.

The stepwise multiple linear regression analyses conducted to test the hypothesis that "sets of the intra-oral variables

measured have a significant capacity to predict or explain variation in individual caries experience", were significant at the  $P < 0.001$  level. With the presentation of these results the second specific aim of the investigation has been achieved.

The highest multiple correlation coefficient (0.74) was attained between caries experience expressed as the severity score of carious lesions, and the following independent variables, listed according to the magnitude of their contribution: Colony type 2, Colony type 5, the bicarbonate content of saliva, Colony type 9, the resting pH of plaque, Colony type 10, the viscosity of saliva and Colony type 11. Plaque quantity has contributed indirectly, in that most colony types appeared in the analysis expressed in the form of 'the cultivable number x mg plaque collected'. The correlation coefficient of 0.74, significant at the  $P < 0.001$  level, showed that 54.8 per cent of the total variation in caries experience could be explained by variation in the values of the independent variables listed.

The latter result was obtained in an analysis from which caries-free subjects were excluded. Limiting the analysis to carious subjects resulted in a 16 per cent increase in the proportion of the total variation explained, compared with previous analyses where all subjects were considered. This can be attributed to any one or a combination of the following factors:

- 1) Heterogeneity of the carious and caries-free subjects in terms of the independent variables examined.
- 2) A higher degree of heterogeneity amongst caries-free

subjects than amongst carious subjects.

3) The operation of a variable or variables which have not been measured, modifying the effect of variables included in the analysis and affecting carious and caries-free subjects to a different extent.

The viscosity of saliva has been identified as a source of heterogeneity corresponding to the first two categories. Other specific sources of heterogeneity are of considerable aetiological interest and will be examined in follow-up studies.

The outcome of the discriminant function analyses, used at this stage as an exploratory procedure, surpassed all expectations. As in other analyses, variables associated with the quantity, properties and microbial composition of plaque produced the best discrimination. Using this group of variables, it was possible to classify 89 out of 93 subjects who were either caries-free or had six or more DMFT, into their respective caries prevalence categories, on the basis of the information contained in the independent variables alone. Thus, it has been established that the data collected can be used to classify subjects with contrasting caries experience, at an error rate of 4.3 per cent, and the third specific aim of the investigation has been achieved. Misclassification of the four subjects may have been due to 'false error' or to a self-imposed rule of using equal positive and negative error rates. (Section 4.5.)

Variables related to the properties and composition of enamel and saliva, and a combined set of variables, produced a

misclassification rate of 18.0 per cent.

The generally high rate of success of these analyses is attributed to the facts that, in discriminant function analysis, linearity of relationships is not a requirement as in regression analyses, and that contrasts between independent variables are also taken into consideration.

Exploration of further applications of discriminant function analysis, including some which may be capable of interpretation in a casual context is strongly warranted.

Overall, the data evaluating process has not been used to its full capacity, but its comprehensive nature and versatility have been demonstrated.

Computer processing of the information is almost essential in a study of this magnitude. However, it has been shown throughout the text that close personal examination of details of the data cannot be neglected, even when a comprehensive data processing system is used.

The aetiology of dental caries is multi-factorial and its investigation demands a multi-directional approach. The study presented in this thesis demonstrates the feasibility of designing, executing and evaluating investigations based on this principle.

### 5.3. Summary.

Methods employed for the collection and processing of materials and data representing intra-oral and environmental

variables, in respect of a primitive population exhibiting contrasting caries prevalence, were presented.

Data summary tables, showing the mean values and standard deviations for each of the 114 variables examined, by sex, place of residence and caries prevalence level, were documented.

A correlation matrix was constructed which shows the product-moment correlation coefficients, significance levels and the number of observations available for each pair of variables.

Stepwise multiple linear regression analyses were employed to examine the collective capacity of sets of independent variables to explain variation in caries experience. The highest multiple correlation coefficient ( $R = 0.74$ ) was attained between a weighted expression of caries experience, the severity score of carious lesions, and a selected set of ten independent variables. The correlation coefficient of 0.74 is equivalent to the explanation of 54.8 per cent of the total variation in caries experience, by variation in the values of the independent variables.

Discriminant function analyses showed that it was possible to classify subjects of contrasting caries experience (0 DMFT and 6 or more DMFT) into corresponding groups, on the basis of the information contained in sets of independent variables, with a rate of misclassification ranging from 4.3 - 18.0 per cent.

The aetiological, biological and statistical implications of the findings were discussed.