高血压联盟讨论问题

数据采集

* **确定未来要采集的内容；**

建立数据库需要确定字段，这个内容如何确定；

* 历史数据是否全部保留；

历史数据部分冗余和不规范，需要作进一步筛选和整理；

* 整合字段（结合历史数据和未来数据）——建数据库；

历史数据是单独整理还是 新旧数据整合在一起；

* 采集方式确定；

如何采集对应的数据？

* 采集数据入库（前结构化和后结构化）

如果采集的数据是非机构化或者是半结构化对的就需要做处理才能入库；结构化方式有哪些？

数据处理

* 数据分析；

未来获取数据之后如何利用，需要计算出哪些指标结果；

* 数据导出；

数据是否需要导出？导出格式如何？导出内容哪些？

1、将已经整理出的字典表给王教授看了，王教授说从贵阳回来后仔细看看，并找其他几个老师开会讨论，再确定哪些字段需要，哪些不需要。

2、每次随访的问卷，都是不一样的，可能没办法完全兼容以前的数据，并且成本可能会很高。这个也给王教授说了，王教授说可以考虑只兼容共性的字段。

3、随访、突发事件，可以参考之前给我们的那几个英文模版。

4、数据导出这一块，二维表导成一维表的难度，也给王教授说了，王教授说开会的时候问问其他几个老师，看他们有没有好的意见。