Clinical Evaluation of Isoflurane

STUDY DESIGN AND ANALYSIS

CHARLES H. GOLDSMITH

The purpose of this paper is to outline the methodological considerations inherent in conducting a large scale study to investigate factors related to ways of handling a new anaesthetic agent – isoflurane.

The objective of the study was to investigate the relationship (if any) of isoflurane concentrations and delivery characteristics to demographic characteristics of patients, drugs administered simultaneously, pulse rate, systolic blood pressure, surgical procedure, muscle relaxants, reflex activity, arrhythmias, and any complications that the attending anaesthetist considered to be important.

STUDY DESIGN

The design which the authors agreed on was a single group cohort where all eligible patients would be given a suitable dose of isoflurane. It was hoped to involve about 200 North American teaching hospitals and that each of these hospitals would contribute up to 100 patients to the study between 1980:11 and 1981:03. Compliance with this guideline would have generated a maximum of 20,000 patients who would have had isoflurane as an anaesthetic agent and extensive data would have been generated for relationship determination and possible adverse reactions. A standardized two-sided data collection form was completed on each patient entered into the study. The variables measured were partitioned into those events monitored during preinduction, induction, maintenance and postoperative periods. (See Appendix I).

Inclusion criteria consisted of patients of all ages, gender, and ASA status. Obstetrical patients were specifically excluded.

Using the rule of three[42], with 95 per cent confidence, adverse reactions with an incidence of less than one in approximately 6,700 (<0.0002) would be detectable with such a design size.

Once each form was completed for a patient, the form was sent to the monitor center who checked over the information completed on the form. When the monitor was satisfied that the information was complete, the form was forwarded for data analysis.

ACTUAL DESIGN

Table 44 shows the summary information for the actual design. There were some 165 centers contributing 7,196 patients of which 6,798 (94 per cent) were eventually entered into the data base for subsequent analysis. Only 7/165 (4 per cent) of the centers contributed at least 100 patients. Table 44 also shows the number of centers in each of the regions in Canada and the United States as well as the number of cases contributed by each region. The median number of cases per center was 35. The standard deviations are all of the order of 30, indicating that there was great variability in case number contributions from hospital to hospital within each of the regions.

DATA FLOW

Figure 11 outlines the data flow characteristics from patient consideration through to final analysis. A patient was considered for eligibility, and if they were not eligible, they were excluded. If they were eligible, they were anaesthetized with isoflurane and the form (Appendix I) was completed. The monitor at the center would check out the form, and if the data appeared to be satisfactory, the form was sent to the Madison Center where a central check was done on all of the forms coming from all of the reporting hospitals. If either the monitor or the Madison Center found some anomaly, the form was returned to be corrected. Once the forms had passed the Madison Center checks, they were forwarded to the New Jersey Center where they were keypunched and verified and eventually entered into a Hewlett-Packard (HP) 3000 com-

TABLE 44

NUMBER OF HOSPITALS AND NUMBER OF CASES PER HOSPITAL BY REGION

Region	n	# > 100	Cases					
			Total	Minimum	Median	Mean	Maximum	s
Canada	35	3	1829	1	48	52.3	151	37.0
United States								
North Eastern	52	0	1707	7	28.5	32.8	88	20.5
North Central	32	2	1417	4	37.5	44.3	183	37.2
Southern	24	1	967	8	38	40.3	122	23.2
Western	22	1	877	5	29.5	39.9	149	31.5
Total	165	7	6798	1	35	41.2	183	30.5

n = Number of hospitals
# > 100 = Number of hospitals in region contributing at least 100 cases
s = Standard deviation

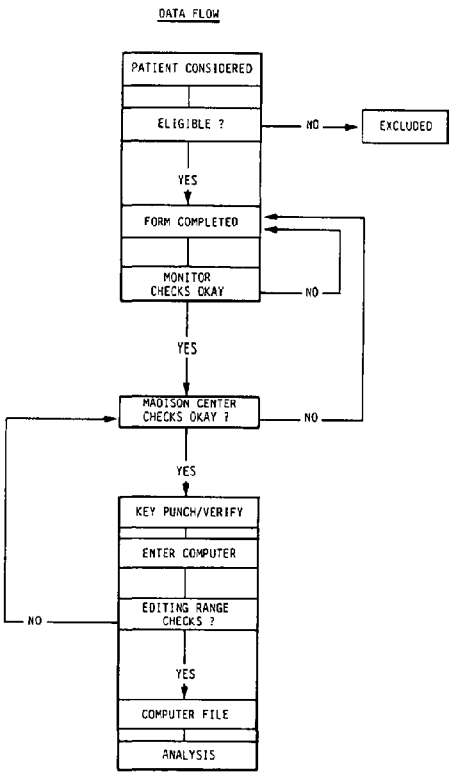

FIGURE 11 Flow diagram of process involving scrutiny and logic checks for data collection forms prior to entry in the computer for complete data analysis. Initial screening was done at the Madison Centre while range checks and analysis was done at the New Jersey Centre.

puter where a set of edits was completed on each record to make sure that each piece of information satisfied specific range criteria. If errors were found, they were sent back to the Madison Center for clarification and the file was not loaded until the data passed the range check. Once the range checks had been verified, the data were put on a computer file and eventually used in the analysis.

DATA QUALITY

The range checks to which the data were submitted included cross references for age, weight, height, pulse, and blood pressure, isoflurane concentrations and fresh gas flows. Additional checks were made on muscle relaxant doses.

METHODS OF ANALYSIS

After the data had been successfully edited and stored on the computer file, the BMDP[43] and SPSS[44] statistical analysis packages were used to form summary descriptive statistics, scattergrams, correlations, simple regression analyses, t-tests, analyses of variance, chi squared tests, log linear models, logistic regressions and multiple linear regressions.

Summary data were provided on an ongoing basis to the panel to provide for feedback and to check the structure of the analyses. In addition, a preliminary analysis based on a sample size of 2,027 cases, available at the time, was summarized and presented to representatives from each of the contributing hospitals.

The final analysis is reported for 6,798 cases on the computer file as of 1981:04:15.

Because each of the outcomes was likely to be challenged by multiple (k) subgroups or independent variables, the conventional level of significance, α, was modified to a more stringent level of significance, α/k. In the reports, results have p values that are rounded to four decimal places and their interpretation is as follows: 1) declare the relationship to be *not-significant* if it happens to have a p value greater than $\alpha = 0.05$; 2) declare the relationship to be interesting and *worthy of further study* if the p value lies between $\alpha = 0.05$ and $0.05/k$; and 3) declare the relationship to be *significant* if the p value is less than $0.05/k$.[44] The number of challenges would be defined for each outcome separately.

Results

Learning Curve

Since all participants were using isoflurane for the first time, it was postulated that a learning phenomenon might occur and that those centers that would only deal with a small number of cases might have much higher and more variable levels of isoflurane during induction and maintenance. Subsequently, using the SPSS[44] REGRESSION Subprogram, the simple regression model of the isoflurane variable on the number of cases dealt with by a hospital was computed. Table 45 shows the slopes and the p values for average induction dose, highest induction dose, average maintenance dose, and highest maintenance dose. None of the p values came close to being less than 0.05. We conclude that there was no relationship between each of those concentrations and the number of cases submitted per hospital. In other words no learning curve was present in the data in terms of concentrations used. However successive cohort analysis did show a trend to higher ASA status, higher age, more major procedures with time and a slightly lower incidence of reflex activities.

TABLE 45

Simple Linear Regression of Per Cent Isoflurane Measurements by Number of Cases Per Hospital

Variable	Slope	p
Average Induction Dose	0.0017	0.2717
Highest Induction Dose	0.0014	0.1720
Average Maintenance Dose	0.0008	0.2530
Highest Maintenance Dose	0.0005	0.2418

All p values based on 1,163 df from an F statistic

Discussion

Some 6,798 cases were admitted to the study and analysed. If no adverse reaction was observed in this group of patients, then with a probability of 0.95, the incidence of that adverse reaction is likely to be less than one in approximately 2,300 (<0.0004). This incidence rate happens to be approximately twice what was hoped for in the original design. Consequently, if one believes that all of these cases were collected with reasonable comparability, then only extremely rare adverse reactions would be missed by this study.

Although each center received a calibrated vaporizer specific for isoflurane and they received sufficient isoflurane to anaesthetize approximately 100 patients, only seven of 165 centers were able to generate 100 or more patients.

Of the five per cent of patients entered in the patient log but not included in the data base, only incomplete information was available. These patients, examined globally and specifically for reflex activities, arrhythmias and complications, were found to be no different than the complete cohort of 6,798 patients.

The completion of this single group study and the findings reported here should be compared with historical controls at a minimum and compared in a prospective randomized controlled trial in the ideal. Thus the clinical effects of isoflurane can now be compared to other anaesthetic agents such as enflurane and halothane using common data collection techniques.

One also has to remember that in surveillance studies such a this one, "... if you follow all people in the world to the end of time, you are going to find every adverse reaction there is – whether or not they are taking drugs. So I think that we have to be careful in our judgment as to when, where, and how we are to look".[46] In a single group cohort design, we have to recognize possible specification and selection biases.[47] For example, the *popularity bias* could lead to admission of patients to these university hospitals because of the interest stirred up by this anaesthetic agent and/or other presenting conditions. The *referral filter bias* might mean that the university centers could be the recipient of the

more complicated and multiple diseased patient, whereas some of the primary and secondary facilities might have less complicated disease referral and could have a slightly different picture with respect to how isoflurane is handled. The *missing clinical data bias* has already been alluded to and, as far as we can deduce, this is not an important consideration. The *non-contemporaneous control bias* has already been alluded to because this study was a single group cohort. This limits the comparison of isoflurane to previous experiences using quite different data collection forms.

There may also be biases in the use of isoflurane as the anaesthetic agent. For example, the *contamination bias* is prevalent throughout this study since the anaesthetist who administered the anaesthetic agent had the opportunity to alter flow rates, concentrations, other medications, etc., and very likely did so; and consequently we are unable to state the cause of any specific outcome with a degree of certainty. The *therapeutic personality bias* may also be present since many of the people (including monitors) filling out the forms may have desired positive results from this anaesthetic agent and hence may have inadvertantly modified the results that were recorded.

Biases that can enter the data analysis have been identified and appropriate precautions have been taken. The *data dredging bias* has been accounted for by adjusting the significance level by the number of challenges that have been made of the outcome. No adjustment has been made for the *repeated peaks bias*. However the repeated peaks have not dictated the final analysis in the report.

In the interpretation of the study we may run the risk of having any or all of these biases creep in. One always wonders what is the significance bias. Do we wrongly attribute biological or clinical significance to an event simply because of its statistical significance?

We believe that each significant finding has been suitably tempered with the current knowledge about clinical and biological relationships. Hopefully we have been careful not to infer that demonstration of a significant correlation indicates causation without the ultimate demonstration that we should get a corresponding modification in outcome when we alter the causative agent through a randomized controlled trial.

Although this study has some limitations in design, one should not lose sight of the fact that 165 university teaching centres participated in the clinical evaluation of a new anaesthetic using a standardized form for collecting data. This has provided a large body of information on a broad spectrum of patients presenting for surgery at university hospitals. These data can contribute to our understanding of basic biological mechanisms such as blood pressure, relationships with age, as well as natural history of blood pressure during operation as one example.

SPECIAL ACKNOWLEDGEMENT

The efforts of Mr. Brad Smith deserve special recognition. He was responsible for the computer programming of the entire study. In addition he did additional analyses on arrhythmias, reflex activities and complications which have been incorporated in this report. His painstaking handling of a massive amount of data and the innovative way in which he solved the many extremely complex analytical problems inherent in the study contribute in no small way to the final success of this study.