Observations on the Illinois Lineup Data

Nancy Steblay, Ph.D.
Augsburg College
May 3, 2006

First, Hennepin County --

Over the last two years, I have worked with the Hennepin County (HC) Attorney's Office to analyze data from their year-long pilot project. Today the HC double-blind sequential procedure is regarded as cost-effective and successful. As Illinois decision-makers and others across the country weigh the results of the Illinois project, it is useful to consider the Minnesota experience.

Prior to the start of the pilot program, the HC Attorney's Office was convinced by the scientific laboratory evidence that double-blind sequential procedure was essential to reduce the risk of eyewitness error and to gain greater confidence in eyewitness evidence. Therefore, all lineups in the HC pilot were double-blind sequential. The central question for HC was somewhat different from Illinois: How do blind sequential techniques work in real police investigations? A summary of the Hennepin County project is in press in the Cardozo Public Law, Policy and Ethics Journal[1], and full data will be available later on the website of the National Institute of Justice, which funded my work on the project. The report clarifies misconceptions about the HC protocol that were present in the Illinois summary report.

The aggregate data for 280 HC double-blind sequential lineups:

- Suspect IDs 54 % (of 280 lineups)
- Filler selections 8 %
- No Choice 38 %

Interpretation of these data illustrates two important considerations regarding lineup field data. First, it is essential to be clear about what is being measured; this is particularly true with simple aggregate figures such as the above. Second, the numbers must be viewed in the context of broader considerations of what is known about memory, eyewitness error, and field practice.

Field data may be parsed many ways for analysis, with subsequent effects on the overall aggregate figures that become the simple summary of the more complex data. For example, HC chose to include every lineup conducted for felony crimes in four cities, including crimes in which the perpetrator was a stranger to the witness and those in which the witness was familiar with the offender. Most HC lineups followed crimes of very short duration for which the witness did not know the perpetrator; not surprisingly, this subset resulted in lower suspect ID rates and somewhat higher filler selection rates compared to situations in which the witness had longer or even multiple exposures to the perpetrator. Of course, the aggregate statistics change if the mix of stranger vs. familiar perpetrator percentages that make up the data set is adjusted.

Adherence to lineup protocol is also important. The double-blind sequential procedure is really a package of components that together form best practice. It is useful to keep track of how these components are implemented in the field – and how revisions to the recommended protocol are employed. An example is the use of "laps" in the sequential procedure. In HC, a witness's additional viewings of the sequential lineup generated more known errors (fillers). Would filler rates decline if witnesses were held to a single viewing of the lineup? Now we know: yes. The point is that it is useful to know the particulars of both the witnesses' exposure to the offender and of the identification experience and how these affect the overall statistics.

Now, the Illinois data --

The Illinois cities of Chicago, Evanston, and Joliet collected field data for police lineups under two conditions: 1) double-blind sequential; and 2) traditional non-blind simultaneous ("status quo"). The two conditions provide a picture of how each procedure is operating in the field and may allow a starting point for productive discussion about the effectiveness of field procedures in securing accurate eyewitness memory.

How does the double-blind sequential technique perform in Illinois police investigations?

The prescribed blind sequential protocol for both HC and IL *required* that the lineup include only one suspect. A single-suspect model is the recommended blind sequential procedure, therefore the IL data below are those blind sequential lineups that adhere to that requirement. This also allows for a closer apples-to-apples comparison to HC. Both the Malpass and Ebbeson analyses are presented for IL: Dr. Malpass' data exclude perpetrators known to the witness; Dr. Ebbeson's data include known perpetrators, as do the HC data.

The similarity in performance between IL and MN for these blind sequential lineups is noteworthy. One reasonably may conclude that blind-sequential lineups are producing very good results in both Illinois and Minnesota test sites.

Double-blind sequential lineups; 1 suspect in lineup

Suspect ID	51%	57%	54%
Filler selection	7%	7%	8%
No choice	43%	37%	38%
	(Malpass)	(Ebbesen)	(HC)

How does the traditional lineup perform in Illinois police investigations?

Here also, the witness decision rates suggest reasonable levels of suspect identifications and low filler selections.

Non-blind simultaneous lineups: 1 suspect in lineup

Suspect ID	64%	67%
Filler selection	3%	2%
No choice	34%	31%
	(Malpass)	(Ebbesen)

How do the two Illinois conditions compare?

Before moving to a comparison of the non-blind simultaneous and the double-blind sequential conditions, the nature of the data must be considered.

Eyewitness performance measures.

Witness performance in the field is measured through percentage rates of suspect identifications, filler selections, and "no choice" responses. *Suspect IDs* cannot be directly equated to *Accurate IDs*, because any false identification of an innocent suspect (dangerous error) is buried within the *Suspect ID* category. *Filler selections* are known errors -- not directly dangerous to the selected lineup member, but a signal to investigators that the witness has a poor memory of the offender, or is uncooperative, or perhaps that the filler is a better match to the perpetrator than is the suspect. *No choice responses* include those witnesses unable or unwilling

to select from the lineup; a response of "he's not there" suggests that the suspect is not the offender.

Because these field measures are imperfect, there are no established absolute levels of "goodness," and the evaluation of a procedure must be made within the context of what we know about the data that contributed to the statistical outcome and about estimated gains or loses in accuracy that are likely from the procedures employed.

Lineup Procedure and Construction Issues in the Illinois data

Witness decisions may be affected by both the lineup *procedure* and by the *quality of the lineup construction*. Some of these topics are discussed below.

1. Non-blind lineups

Dr. Gary Wells, in his website comments on the IL report[2], explains the implications of the absence of administrator blind in the simultaneous "status quo" condition of the Illinois pilot, and I refer readers to that document for a more complete discussion. The lack of administrator masking *is* a substantial difficulty in these data and truly does prevent clear understanding of the status quo condition.

Lack of a blind administrator is inherent to the IL "status quo" lineup. That fact, however, should not excuse evaluators from rigorous examination of the role that non-blind administration may have played in the outcome of that field condition. A non-blind administrator may affect the witness's decision and confidence as well as the administrator's attention to and recording of the witness's comments and decision.

Concern regarding non-blind is more than simply a preference for basic scientific rigor. The vulnerability of the traditional lineup to administrator influence and to witness inferences about administrator behavior is a very serious concern and an ever-present challenge to validity of eyewitness evidence. A problem with the data from Chicago and Evanston in particular is that the pattern easily suggests dramatic non-memory influences on eyewitness decisions.

2. Zero filler identifications in Chicago and Evanston non-blind simultaneous lineups

The most surprising element of the IL data is that both Chicago and Evanston report no filler identifications in their non-blind simultaneous lineups (see table on page 5). This suggests a qualitatively different experience from that of simply low filler rates. It is puzzling to see 152 lineups in which no eyewitness chose a filler and to imagine that all witnesses with weak memories and/or limited exposure to the offender were captured in the *no choice* category.

Nearly 3000 data points from simultaneous field lineups have been collected by others[3], showing a typical filler selection rate around 20%. The IL data are substantially at odds with these figures.

3. Muddy areas in the field data

The blind sequential protocol was clearly scripted in the IL study; the requirement for the non-blind simultaneous protocol less so. The practice of the status quo presumably encompassed a flexible range of lineup size and construction, instruction, and presentation modes. Thus, while constraints of the new procedure helped to clarify its parameters, the non-blind simultaneous

condition remained amorphous. We know very little about the status quo condition, beyond Exhibit 13 of the Illinois report.

The IL data analysis tables indicate variation in such factors as lineup size (4 to 46) and number of suspects in the lineup (0-3), as well as differences such as photo vs. live presentation and some instances of multiple lineup tasks for the same witness. These variations extended to both lineup formats.

Some variability in protocol is not a problem, particularly as it may represent natural field practice. The challenge is to evaluate lineup outcomes when procedural changes violate prescribed protocol or when they confound interpretation of the data. For example, placement of more than one suspect in the lineup runs counter to recommended protocol (one reason for my reliance on the *1-suspect* data). Blind sequential protocol is very clear about a-priori definition of one suspect and at least five fillers.

Similarly, understanding of live lineup performance is confused in cases where the witness's selection of the suspect from a photo lineup was followed with a (second) live lineup identification task. This practice introduces the possibility that the witness's recognition of the suspect in the live lineup stems from the previous procedure (rather than the crime) or that the witness can infer the suspect's identity due to his presence as a common denominator across two lineup tasks.

An aside: The IL report compares IL outcomes to Queens (New York) data (pg 43). Those data, from live simultaneous non-blind lineups, show an annual filler selection rate of < 5%. Caution is advised, however, in that these lineups typically are the second identification task for each witness. Eyewitness identification practice in New York, if I understand correctly, usually includes an initial photo lineup task; a successful suspect ID is then followed by the live lineup identification test. It should be noted that Behrman and Davey's 2001 documentation[4] of traditional simultaneous field lineups in California showed that the practice of a using second identification task increased suspect ID rates 17% for the later identification.

Problems of variability are compounded to the extent that they occur disproportionately in one condition versus the other. However, even if a deviation from practice is distributed evenly across both lineup conditions, there may be uneven effects. For example, it may be reasonably argued, following the results of Lindsay, et al., 1991[5], that the sequential format may help to tamp down effects from some aspects of biased lineup construction.

4. Functional lineup size

Functional size of the lineup – the number of lineup members who effectively match the description of the perpetrator -- is an important consideration for both sequential and simultaneous lineups.

When HC results were released, critics reasonably argued that the low filler selection rate (8%) could be due to poor lineup construction. That is, if fillers are implausible, the witness is left with only the suspect. If this were the case, suspect and/or no-choice rates would be elevated and filler rates reduced. HC took this concern very seriously and conducted an extensive mock witness test with a sample of lineups from the original 280. The fillers were found to be functioning well, drawing mock witness responses in appropriate proportions. The alternative interpretation for low filler rate – poor fillers -- can be ruled out in HC. Illinois may wish to conduct a similar test.

5. Inconsistent performance across photo and live lineups.

A comparison of photo to live lineups is provided in Ebbesen's Table 8. Note that for both simultaneous and sequential lineups in Chicago, the suspect ID rate is higher for physical lineups. However, In Evanston, suspect ID rate is lower for physical lineups. (No live lineups were used in Joliet.) The reasons for asynchrony in Chicago vs. Evanston and live vs. photo are not evident. However, one can speculate that the quality of fillers in live vs. photo lineups differs or that the procedure in live vs. photo arrays somehow varies, even across jurisdiction.

	Chicago		Evanston	
	Nonblind Sim	Blind Seq	Nonblind Sim	Blind Seq
Photo Suspect ID	63%	49%	76%	47%
Live Suspect ID	76%	61%	60%	33%

6. Inconsistent performance across jurisdictions

Joliet produced positive performance for blind-sequential Suspect IDs. (This outcome was mis-stated in the Illinois report (p. iv) where it was claimed that all three cities showed inferior performance by the new procedure.) Joliet represents a full 33% of the lineups. It seems important to determine why Joliet results were so different from the other two cities. Joliet ran only photo lineups – might this be one reason for the difference?

	Chicago		Evanston		Joliet	
	Nonblind Sim	Blind Seq	Nonblind Sim	Blind Seq	Nonblind Sim	Blind Seq
Suspect	64%	50%	72%	44%	62%	69%
Filler	0%	6%	0%	14%	4%	7%
No choice	36%	44%	28%	42%	34%	24%

(Data from Ebbeson report, pg 7)

The Malpass data in Table 3b in fact provide a strong case that Joliet should continue to use double-blind sequential lineups.

Joliet	Nonblind Sim	Blind Seq
Suspect ID	61%	69%
Filler	6%	3%
No choice	34%	29%

7. Laps

The HC data showed that additional viewings (laps) of the sequential lineup reduced eyewitness accuracy, and this lesson was repeated in the IL data. The simultaneous lineup format does not allow us to know the level of comparison shopping (relative judgment) employed by the witness prior to his or her decision. In the sequential lineup, the number of laps is an objective indicator of the witness's scanning of the lineup. The Malpass analysis shows a 5% higher suspect ID rate for the "one lap" witnesses, bringing suspect IDs in that group to 50%. Aggregate data that include witness decisions following multiple laps will present a less positive picture of the double-blind procedure.

The broader context for data evaluation – A reminder about eyewitness memory

The objective of a lineup is to accurately access the eyewitness's memory of the perpetrator, circumventing non-memory factors that may otherwise influence the witness's decision. Examples of non-memory factors include expectations and beliefs of the witness, inferences that the witness may draw from the lineup construction or presentation, the relative judgment process, and influences (imagined by the witness or real) from the lineup administrator.

The new double-blind sequential lineup procedure was designed to limit the impact of some non-memory factors that may occur in traditional lineups. Much recent laboratory research has occurred in direct response to wrongful conviction data that indicate eyewitness error as a central factor in miscarriages of justice.

Challenges to eyewitness identification evidence typically claim that the witness's decision can be attributed to non-memory factors. Sound eyewitness evidence must be able to demonstrate that the identification is of the highest quality – that extraneous influences have been avoided. Thus, evaluation of field data must include recognition of the extent to which the outcomes represent procedural safeguards against memory error.

So, which technique is better for Illinois?

It is likely apparent at this point that I do not believe a firm answer can come from these data. Any meaningful attempt to compare conditions is made very difficult by the complexities noted above.

My primary concern with the Illinois report is that its conclusion appears to be based primarily on the simple aggregate results in Table 3a, with minimal appreciation of the underlying reasons for these outcomes or the broader context of what is known about eyewitness fallibility.

Illinois decision-makers must determine their next steps from imperfect and somewhat confusing field data. My inclination is to assume that the blind sequential has much to offer and to reject the notion that the status quo should be the "standard to beat," particularly given the demonstrated vulnerability to witness error of standard lineup procedures.

Again, perhaps the HC experience can be useful.

The broader context for data evaluation: What has been gained in Hennepin County?

HC's conclusion is that the blind-sequential procedure is working well in Minnesota. Acceptable suspect ID rates and low filler rates suggest a protocol that will help to convict the guilty and protect the innocent. Beyond the numbers, the following considerations are important in this evaluation:

- HC now has implemented a standardized scientifically-based procedure that works in the field.
- The eyewitness is pushed to depend on memory.
- The lineup administrator conducts and reports an objective appraisal of the lineup interaction and results.
- HC data regarding eyewitness response patterns in the field make sense in the context of what is known about eyewitness experience and memory.
- These results are achieved with the safeguards of the Double-Blind Sequential technique in place.
- HC is now seeing the benefits in court of stronger eyewitness evidence: the new procedure rules out the likelihood of administrator influence, selective or imprecise reporting of filler selections, relative judgment effects, and (with the filler quality test data) ineffective lineup construction.

Footnotes

1. Klobuchar, A., Steblay, N., & Caligiuri, H. (in press). Improving eyewitness Identifications: Hennepin County's Blind Sequential Lineup Project. Cardozo Public Law, Policy, and Ethics Journal.

2. Wells, G. March 29, 2006. Comments on the Illinois Report. http://www.psychology.iastate.edu/faculty/gwells/Illinois_Report.pdf

3. See the following:

Valentine, T., Pickering, A., & Darling, S. (2003) Characteristics of eyewitness identification that predict the outcome of real lineups. Applied Cognitive Psychology, 17, 969-993.

Slater, A. (1994). Identification parades: A scientific Evaluation. Police Research Award Scheme. London: Police Research Group, Home Office.

Wright, D.B., & McDaid, A.T. (1996). Comparing system and estimator variables using data from real lineups. Applied Cognitive Psychology, 10, 75-84.

Behrman, B.W., & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. Law and Human Behavior, 25, 475-491.

4. Behrman, B.W., & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. Law and Human Behavior, 25, 475-491.

5. Lindsay, R.C.L., Lea, J.A., Nosworthy, G.J., Fulford, J.A., Hector, J., LeVan, V., & Seabrook, C. Biased lineups: Sequential presentation reduces the problem. Journal of Applied Psychology, 76, 796-802.

Contact: steblay@augsburg.edu