Limitations of Experimental Human Abuse Liability Studies and Prospects for Improvement

James P. Zacny

Department of Anesthesia & Critical Care

University of Chicago

Outline of Presentation

- Issues relating to human ALA of opioid ADFs

- Issues relating to human ALA of opioids *in general*

ALA= abuse liability assessment
ADF=abuse-deterrent formulation

Issues Relating to Human ALA of Opioid ADFs

- An added component to traditional ALA when testing ADF's: extractability testing

- What is the appropriate subject population for ALA of ADF opioids?

- When doing IN and IV ALA, should the extract from the ADF formulation be used?

Issues Relating to Human ALA of ADF's

- What kind of ALA testing needs to be done with ADFs?

- Differences in latency to peak positive effects of IR vs. ADF ER opioids...how much does it matter?

- What constitutes a clinically significant reduction in abuse liability of an ADF?

An Added Component to Traditional ALA When Testing ADF's

Extractability (Benchtop) Testing

Extractability Testing

- Why do it?
  - We know abusers are going to try to circumvent the abuse deterrent aspect of the formulation
    - To enhance the drug effect (to increase cmax)
    - To achieve a faster onset of effect (to decrease tmax)
    - To alter formulation to allow for different route of administration (from oral to either snorting or iv use)
    - To remove undesirable ingredients (ACET, NLTX)

Major Issues Related To Extractability Testing

- What materials are needed for extraction studies?
    - Crushing equipment; solutions/chemicals; heating/freezing equipment
- From an abuse deterrent perspective…
    - The more materials needed and the more esoteric they are in order to extract, the better the ADF
    - The more complicated the extraction process is and the more time consuming, the better the ADF

Major Issues Related To Extractability Testing

- How much of the opioid can one extract from the ADF?
  - Example: 120 mg MOR ADF ER
    - 20 mg (good ADF) vs. 100 mg (hmmm…)
- With ADFs that have other products in them that are undesirable to the abuser, how much of those products can be extracted (removed)?
    - Examples: aversives (blockers, niacin, capsaisin; acetaminophen, talc/ binders)
   - More aversives that can be removed, more likely that the ADF will be snorted or injected

Extractability (Benchtop) Testing: The Problem

- "Neither the FDA nor the DEA have objective criteria to measure extractability."

- "Standardized laboratory extraction must be developed for each type of formulation (e.g., tablet, capsules, patches) using solvents and equipment commercially available."

Extractability Testing: A Template

Taylor & Francis
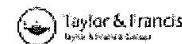
Development and Preliminary Experience with an Ease of Extractability Rating System for Prescription Opioids

N. P. Katz
Inflexxion, Inc., Newton, MA
and Tufts University School of
Medicine, Boston, MA

D. C. Buse
Harvard Medical School,
Boston, MA

ABSTRACT One important factor in the abuse potential of an opioid product is the ease with which active drug can be extracted. There are currently no standards for testing or reporting extractability. This article describes the development of an Extractability Rating System for use by the pharmaceutical industry and regulators. Despite several limitations, this effort serves as a call

Katz and colleagues designed a system to measure the ease/ difficulty by which the opioid can be extracted from the formulation (e.g., ADF). (Katz et al., 2006 Drug Develop Ind Pharmacy 32:727-746)

However, to date there appear to be no industry-wide standards on extractability testing.

2. What is the appropriate subject population for ALA of ADF opioids?

- Consensus in ALA community is that subjects…
  - must be recreational users of prescription opioids
  - must be able to detect an effect of an opioid with known abuse liability, and show a positive response to that opioid
  - if going to be testing IN or IV route, the user should have a history of such use
- However, no criteria on usage history

2. What is the appropriate subject population?

Drug	Age	Number of times used recently	Number of times used in last year	DSM-IV Opioid Use Disorder	Prescreen for liking
IR with an aversive	18-55	-	-	yes	no
ER opioid A - ADF	18-55	At least once in last 3 months	At least 10 times	no	no
ER opioid A - ADF	18-50	-	At least 5 times	no	yes
ER opioid B - ADF	18-65	At least once in last 30 days	-	no	yes
IR opioids	? (28.6)	? (1-2X weekly)	?	?	not clear

2. What is the appropriate subject population?

- Different criteria on use (recent or last year) <u>problematic</u> in studies designed to test abuse liability in nonmedical users
  - As an example…
    - Two studies examine abuse liability of Opioid X
      - one study looks at users with a light history (5X/yr)
      - another study looks at users with a heavier history (1-2X/wk)
      - Ensuing result: possible variability in key endpoints and different interpretations of abuse liability of Opioid X

2. What is the appropriate subject population for ALA of prescription opioids?

- Heroin abusers?
  - Not for testing of oral opioids that are likely to be abused orally
    - Heroin abusers might rate their liking relative to how much they like iv heroin (Fraser et al., 1961, 1978)
    - Increases risk of a Type II error
  - Possibly for studies which assess abuse liability of an ADF via the iv route
    - In fact since the majority of POAs swallow intact or chew*, might be hard to do such an ALA with POAs

* Katz et al. 2008 Internet-based survey of nonmedical prescription opioid use in the US. Clin J Pain 24:528-535

3. When doing IN and IV ALA, should the extract from the ADF formulation be used?

Risk identification, risk assessment, and risk management of abusable drug formulations☆

Curtis Wright IV, E. Douglas Kramer, Mary-Ann Zalman,
Meredith Y. Smith, J. David Haddox *

Risk Management and Health Policy, Purdue Pharma L.P., One Stamford Forum, Stamford, CT 06901-3431, USA

Abstract

There is a demand for pharmaceutical products with reduced abuse liability. These products must meet three tests to be successful. They must be safe for patients, be less likely to injure the abuser, and be less desirable for abuse by established drug abusers relative to existing products on a dose for dose (milligram-equivalent) basis. There is a need for standardization of the evaluation of abusable pharmaceuticals in the various stages of drug development from preclinical animal studies to postmarketing surveillance. Formulations with reduced abuse liability must: (1) be tested using standard animal, benchtop, and human pharmacokinetic methods that allow interpretation, (2) sufficiently reduce the recovery of abusable drug substance, or contain another ingredient to deter abuse, (3) not alter drug activity for patients in an undesirable or risky way, and (4) have an accurate pre-approval estimation of their reduced abuse liability, which is validated by adequate epidemiologic post-approval surveillance.
© 2006 Elsevier Ireland Ltd. All rights reserved.

Keywords: Abuse potential; Prescription drug abuse; Regulation; Medical use

1. Introduction

Drug abuse involving medicinal products has been part of the American experience for well over a century (Musto, 2002). Periods of increased medical use of a drug product or class of products with abuse potential are usually followed by reports of upturn in both drug experimentation among the adolescent population in the United States (cannabis and Ecstacy (3-4-methylenedioxymethamphetamine, MDMA)) (Substance Abuse and Mental Health Services Administration (SAMHSA), 2003a) and in the worldwide increase in illicit production since the mid-1980s of opium and coca, the starting materials for

3. When doing IN and IV ALA, should the extract from the ADF formulation be used?

- Intranasal
  - How safe is insufflating ground drug product? (i.e., does it have excipients that could be harmful to subject?)
- IV
  - Does one try and filter and inject extracted drug after heating it in a teaspoon, or…
  - Does one administer a suitable parenteral dosage based on recovery from benchtop tampering studies?

Wright et al. 2006 Drug Alcohol Depend 83S:S68-S76

3. When doing IN and IV ALA, should the extract from the ADF formulation be used?

- Example of a potential IV ALA study:
    - tampering produces a less-than-pure substance (e.g., solution is slightly viscous or contains particulates but can still load into a syringe)
    - it would be important to know if iv prescription opioid abusers would still like the product
    - questionable as to whether this could be ethically tested
- "Evaluating these issues…is subject to practical and ethical constraints that are always difficult, and often present insurmountable barriers to accomplishment." (Wright et al., 2006)

4. What kind of ALA testing needs to be done with ADFs?

Research design strategies to evaluate the impact of formulations on abuse liability

Shelley McColl [a], Edward M. Sellers [a,b,*]

[a] Ventana Clinical Research Corporation, Toronto, Canada
[b] Departments of Pharmacology, Medicine and Psychiatry, University of Toronto, Toronto, Canada

Abstract

Scheduling of a chemical drug substance under the Controlled Substances Act (CSA) includes an evaluation of preclinical and clinical safety, and experimental abuse liability studies, as well as information on diversion and overdose. Formulations that mitigate abuse liability, dependence potential and public health risks (e.g., altered absorption rate and tamperability, long half-life, pro-drugs and combination products) are amenable to preclinical and clinical studies to compare their abuse potential to reference compounds. For new formulations (NF) as marketed agents, direct comparison to the immediate release (IR) formulation of the reference compound is typically needed across the full range of potential studies.

While the public health advantage of formulation changes in the marketplace can be conceptualized in behavioral economic terms, generating

4. What kind of ALA testing needs to be done with ADFs?

- Conditions
  - The intact ADF (2-3 doses)
  - The IR that contains the same opioid in ADF, in order to show reduced liability (1-2 doses)
  - Placebo
  - In <u>some</u> cases the tampered ADF
    - If opioid is likely to be crushed, then a crushed product
    - Gets dicier with intranasal and iv routes

5. Latency to peak positive effects of IR vs. ADF ER opioids...how much does it matter?

– According to the <u>rate-of-onset hypothesis</u>, drugs that hit the brain faster and produce peak effects soon after administration are more "rewarding" than those that produce peak effects after some prolonged period of time.

– "… issue that deserves consideration is the weighting of the early (e.g., up to 3 h post-dose) and later subjective reports for IR compared to ER formulations". (McColl and Sellers, 2006)

– Should there be a cutoff (e.g., within 3 hours post-drug) in which abuse-liability related effects are given more weight? A "critical window"…?
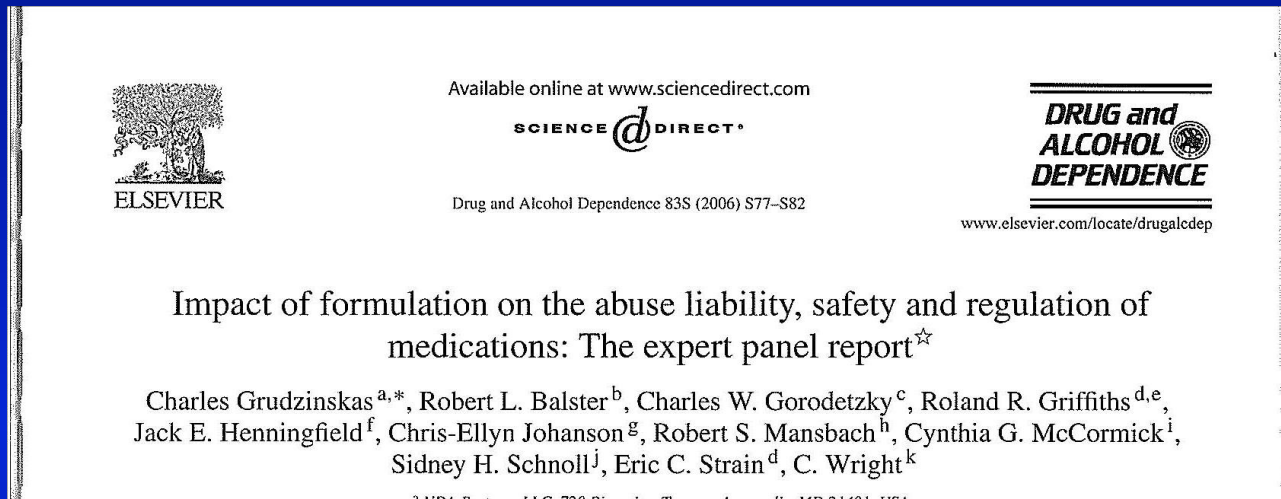
5. Latency to peak positive effects of IR vs. ER opioids...how much does it matter?

- A study that illustrates the weighting issue:
  - In a published abstract the following formulations and doses were tested:
    - 8 mg IR opioid (positive control)
    - 8 mg (crushed), 16, 32, and 64 mg of ADF ER opioid
    - placebo
  - Results
    - All doses and formulations produced higher liking ratings than placebo.
    - Maximum drug liking for the 32- and 64-mg ADF ER was *similar* to 8-mg IR, but occurred later (15 h vs. 2 h).

5. Latency to peak positive effects of IR vs. ER opioids...how much does it matter?

- So... the ER formulation at high doses had equivalent liking ratings as a lower dose of an IR formulation <u>albeit at a later time point</u>.

- How much should the delayed liking of the ER formulation be "discounted"?
 - Should that ER formulation be considered an ADF?

The issue of weighting due to rate of onset not cut-and-dried…

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Drug and Alcohol Dependence 83S (2006) S77–S82

DRUG and ALCOHOL DEPENDENCE

www.elsevier.com/locate/drugalcdep

Impact of formulation on the abuse liability, safety and regulation of medications: The expert panel report☆

Charles Grudzinskas [a,*], Robert L. Balster [b], Charles W. Gorodetzky [c], Roland R. Griffiths [d,e], Jack E. Henningfield [f], Chris-Ellyn Johanson [g], Robert S. Mansbach [h], Cynthia G. McCormick [i], Sidney H. Schnoll [j], Eric C. Strain [d], C. Wright [k]

"However, despite this generally accepted theoretical framework *(i.e., rate of onset)*, the empirical database is limited and the predictive validity of changes seen in well-controlled experimental studies examining differences in rates of absorption to post-marketing deterrence of abuse is inadequately characterized."
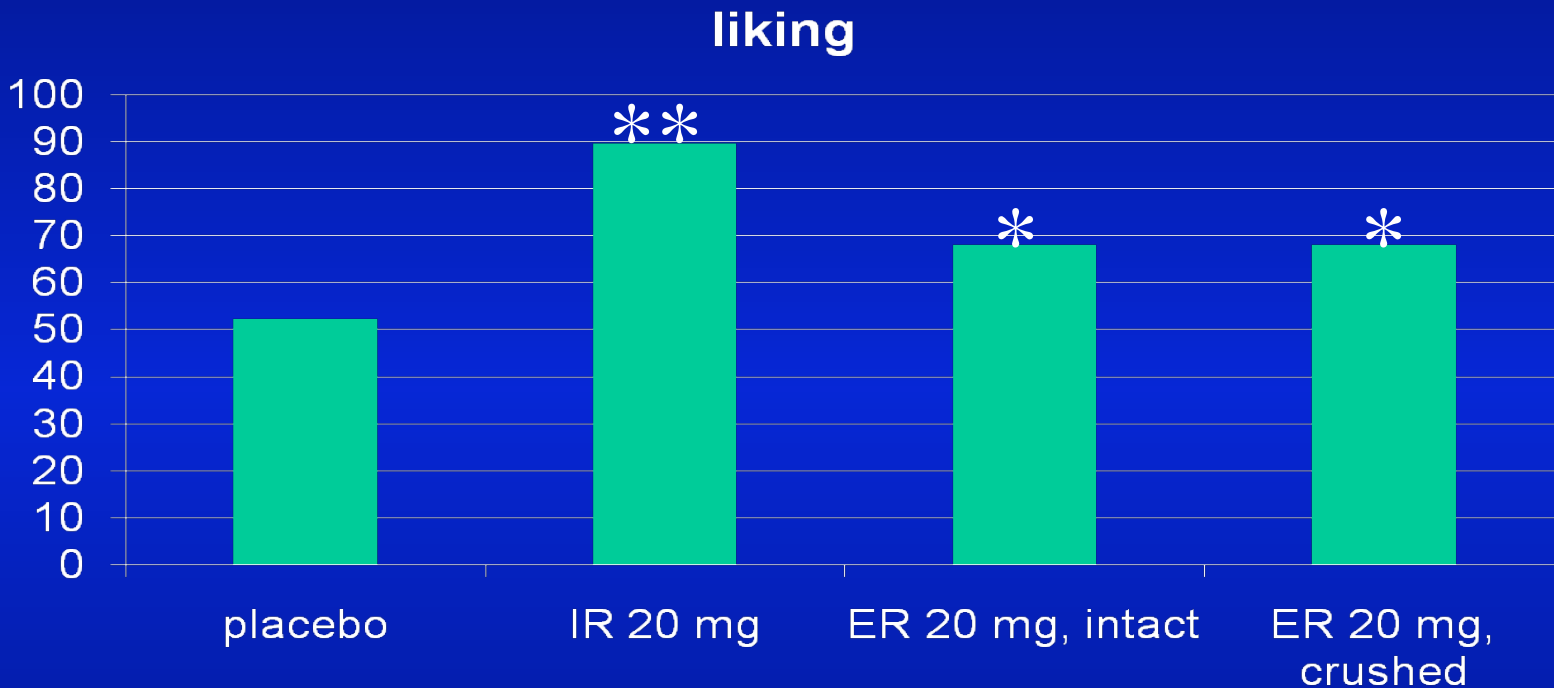
6. What constitutes a *clinically significant* reduction in abuse liability of an ADF to warrant a specific mention in the PI?

6. What constitutes a *clinically significant* reduction in abuse liability of an ADF?
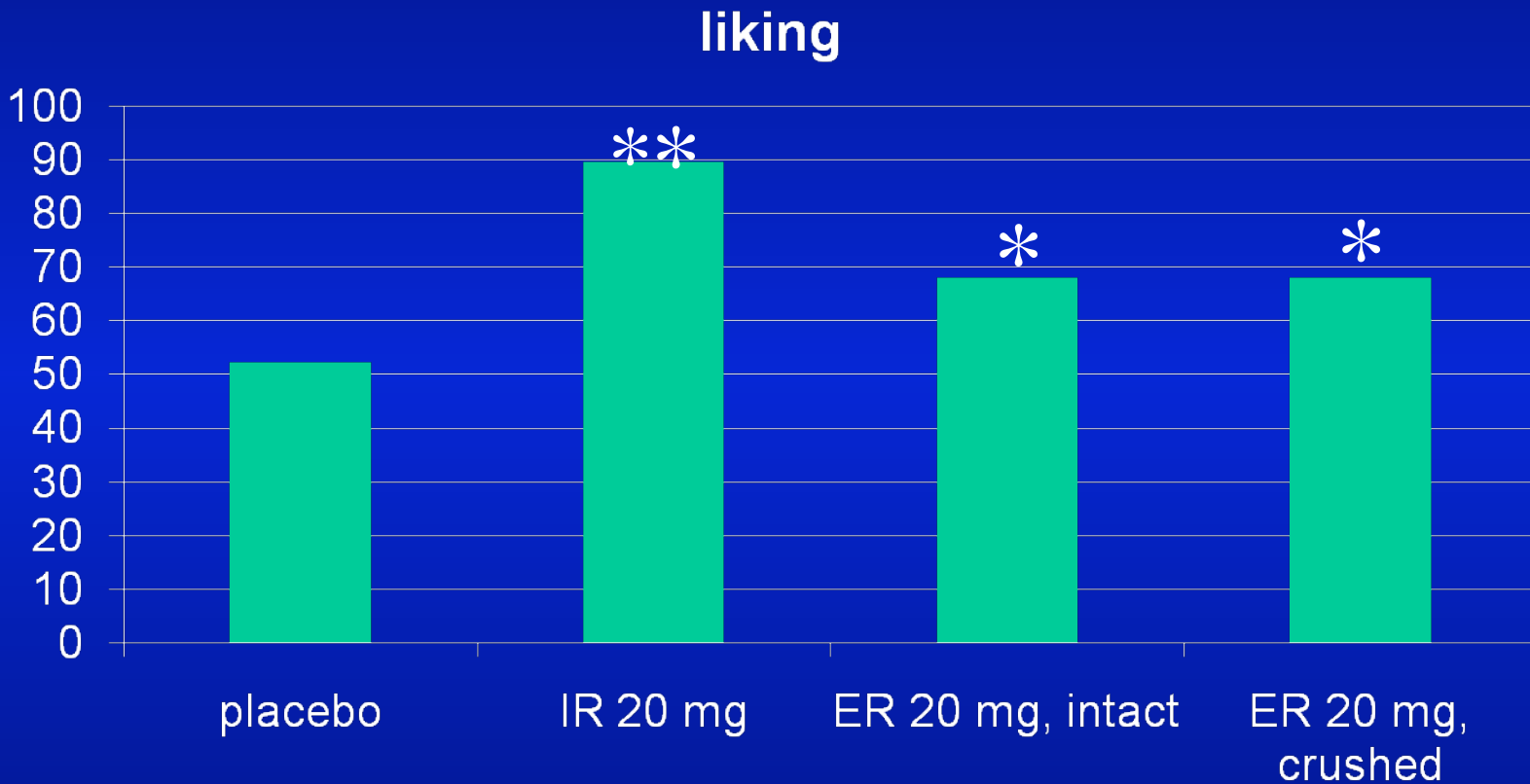
liking

0=dislike a lot; 50=neutral; 100=like a lot

6. What constitutes a *clinically significant* reduction in abuse liability of an ADF?

liking

How does one evaluate the "lowered" abuse liability of the ER product when it produces more liking than placebo? Maybe taking two of the ER tablets would elevate liking to the same degree as the IR product.

6. What constitutes a *clinically significant* reduction in abuse liability of an ADF?

subjective drug value [$]
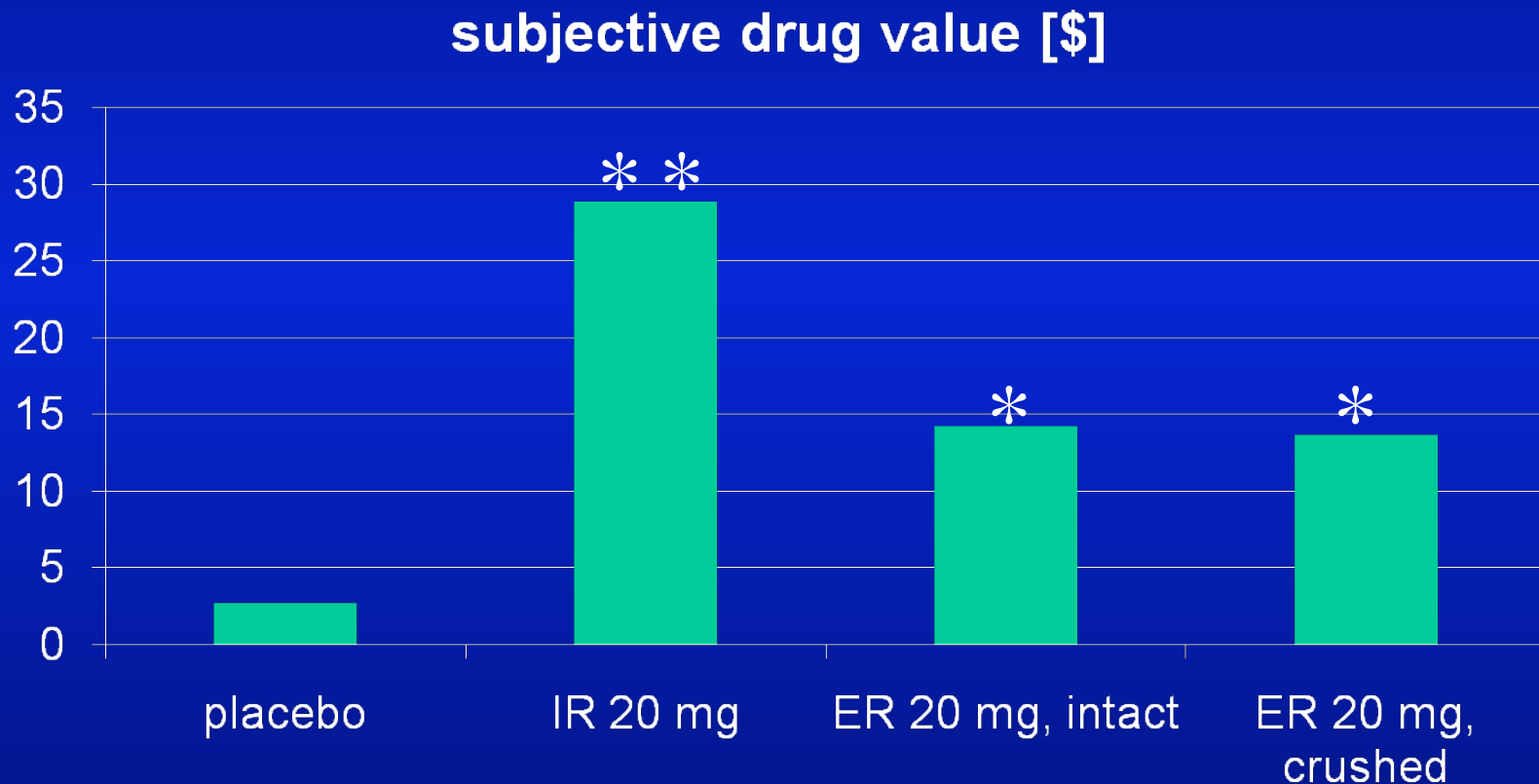
Some issues that relate to human opioid ALA *in general*

1. Does ALA have predictive validity?

2. Have ALA instruments been validated?

3. Is there the need for, or would it be desirable, to standardize primary outcome measures in ALA?

Does ALA have predictive validity?

- Does it predict real world abuse?
  - Well…. yes and maybe
  - Yes
    - If ALA study reveals that Drug X shows no signs of abuse potential in abusers, extremely unlikely the drug will be abused in the world (e.g., diphenhydramine)
  - Maybe
    - If an abuse liability study reveals Drug Y shows strong signals of abuse potential in abusers, the drug may be abused in the real world.

Does ALA have predictive validity?

- The concept of "necessary but not sufficient"
- An opioid must show signals indicative of abuse liability for it to be considered as a potential drug of abuse in the real world.
    - But it is not SUFFICIENT in predicting real world abuse
- Other factors play important roles in actual abuse of an opioid:
    - Its availability (if not readily available, less abuse)
    - Its cost (in and of itself and relative other opioids)
    - What other peers are abusing (e.g., drug d'jour)
    - Potency (low potency, low abuse)

Have ALA instruments been validated?

- IN ALA, there are multiple ways to measure abuse liability
  - Most subjective measures that do not necessarily measure the same thing show concordance on abuse liability of an opioid
    - e.g., MBG (Euphoria), Liking and Take Again
  - This would seem to argue typical measures of abuse liability possess construct validity
  - However, novel measures to assess abuse liability should be subjected to validation testing

Is there the need to standardize primary outcome measures in opioid ALA?

CPDD Expert Panels recommendation

- 2003 (ALA of CNS Drugs)
  - "Researchers should be encouraged to standardize some psychometric scales (e.g. drug liking scales) for human ALA studies, in order to facilitate comparisons across ALA across research laboratories and across drugs."

- 2006 (Drug Formulation and Abuse Liability)
  - "Continued standardization of the primary outcome measures used in abuse potential studies should be encouraged to permit more systematic comparisons across studies."

So what should *the* primary outcome measure be?

- In review articles of ALA, <u>drug liking</u> is considered one of the most, if not the most, sensitive and reliable measure of abuse liability

- Current "problem" is that liking is measured in abuse liability studies in a number of different ways

Drug liking

- Bipolar
  - VAS: 0= dislike a lot; 50=neutral; 100=like a lot
  - Discrete scale:
    - 1=dislike a lot; 4=neutral; 7=like a lot
    - -4=dislike a lot; 0=neutral; +4=like a lot
- Unipolar
  - VAS: 0=not at all; 100=extremely
  - Discrete scale: 0=not at all; 1= a little bit; 2=moderately; 3=quite a bit; 4=very much

Issues Discussed:
A number of them unresolved…

- An added component to traditional ALA when testing ADF's: extractability testing
- What is the appropriate subject population?
- When doing IN and IV ALA, should the extract from the ADF formulation be used?
- What kind of ALA testing needs to be done with ADFs?
- Latency to peak positive effects of IR vs. ER opioids...how much does it matter?
- What constitutes a clinically significant reduction in abuse liability of an ADF?
- Does ALA have predictive validity?
- Have ALA instruments been validated?
- Is there the need to standardize primary outcome measures in ALA?

Acknowledgments

- Kenzie Preston, Ph.D.
- George Bigelow, Ph.D.
- Eric Strain, M.D.
- Craig Rush, Ph.D.
- Lynn Webster, M.D.
- Sandra Comer, Ph.D.